A Solution State NMR Study of the Structure and Ligand Binding Properties of the Human C-type Lectin DC-SIGNR.

Fay Probert

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy.

Molecular Organisation and Assembly in Cells Doctoral Training Centre

University of Warwick

December 2012

Table of Contents.

List of Figures.

List of Tables.

Acknowledgements.

Firstly, I would like to thank Dr Ann Dixon and Dr Dan Mitchell for their supervision and guidance throughout the course of this work and for giving me the opportunity to take up this interesting project. Their support, suggestions and direction have proved invaluable in enabling me to complete this research.

I would also like to thank Dr Sara Whittaker for many extremely helpful discussions, for taking the time to train and supervise me on the spectrometers at the HWB-NMR facility and providing me with 800 MHz spectrometer time. Sara has always shown interest in my work and is always there with helpful advice for which I am extremely grateful. Thanks go to everyone at the Birmingham University NMR facility for making me feel welcome and making my time there thoroughly enjoyable. I would particularly like to thank Dr Tim Knowles for his advice regarding structure calculations using the programme CYANA.

Thanks go to Dr Ivan Prokes for taking the time to train me on the 700 MHz spectrometer at Warwick and providing me with spectrometer time.

I would like to thank all members of the Dixon group, past and present, for their help, many useful discussions and for making the office a great place to be. Special thanks go to Gemma Warren, Dharmesh Patel, Mike Chow, Esther Martin, Anthony Nash, and Leo Bowsher for their friendship and support and for the many coffees and hot chocolates we have drank together throughout the course of my project!

I would also like to thank everyone at MOAC, particularly everyone in the 2008 in-take, for all the help and good times shared at secluded hostels and on invaluable team building courses. Special thanks go to Alison Rodger, for giving me a place on the MOAC course.

I cannot describe how grateful I am to my parents, Sharon and Graham Probert, for the love and support they have shown me throughout my life. I could not ask for better parents and they are always at the end of a phone providing advice and reassurance (and listening to me moan without complaint!).

Finally, but by no means least, I would like to thank my amazing boyfriend, Andrew Willis, for his love, support and for putting up with me throughout this project. Thanks for letting me practice my presentations at you, for understanding when I don't have time to do the dishes and for reminding me to come home in the evenings!

Declaration.

The work presented in this thesis is original, and was conducted by the author, unless otherwise stated, under the supervision of Dr Ann Dixon and Dr Daniel Mitchell. It has not previously been presented for another degree.

The work was funded through the EPSRC, through the Molecular Organisation and Assembly in Cells Doctoral Training Centre.

All Sources of information have been acknowledged by means of reference.

Summary.

The protein DC-SIGNR (Dendritic-cell specific ICAM3 grabbing non-integrin related) is a C-type (calcium-dependent) lectin, which binds highly-branched mannose oligosaccharides. DC-SIGNR interacts with a range of deadly diseases via surface glycans on pathogenic glycoproteins, and the ability of DC-SIGNR to increase the rate of infection of viruses including human immunodeficiency virus (HIV) and hepatitis C virus (HCV) makes the study of DC-SIGNR/oligosaccharide interactions very attractive. The research described in this thesis sought to gain insight into the calcium and ligand binding properties of the DC-SIGNR carbohydrate recognition domain (CRD) in solution by utilising solution state nuclear magnetic resonance spectroscopy (NMR).

A protocol for the production of uniformly ^{15}N /^{13}C labelled DC-SIGNR CRD was developed, allowing the acquisition of heteronuclear NMR experiments and the first assignment of the calcium-bound (holo) DC-SIGNR CRD to be reported. The assignment has allowed investigation of calcium and glycan binding, as well as the pH dependence of the DC-SIGNR CRD.

The data presented in this thesis reveal that the DC-SIGNR CRD is highly dynamic in the calcium-free state, with the addition of calcium resulting in global conformational and dynamic changes throughout the CRD. While calcium binding hinders the protein dynamics (particularly in the calcium binding regions), a large degree of mobility remains. The evidence that ligands are released at low pH suggests that DC-SIGNR may act as an endocytic receptor.

In addition to calcium binding, interactions of the DC-SIGNR CRD with a range of ligands were investigated. In particular, interactions with the oligosaccharide $Man_9GlcNAc$ (present on the HIV viral envelope) are described, representing the first direct study of the CRD interacting with a disease-associated ligand. The glycans employed in this study all bind to the primary calcium binding site, supporting previous crystal data. However, each glycan displays distinct patterns of chemical shift perturbations implying that they each have different, extended binding modes. Particularly striking is the difference between the disease-associated $Man_9GlcNAc$ ligand and the ligand present in a previously published crystal structure, $(GlcNAc)_2Man_3$.

An investigation of the dynamics of the CRD in the holo form and bound to the ligand Man_5 shows that the CRD is highly dynamic and that glycan binding further hinders, but does not abolish, the molecular motions. The dynamics data also suggests that a ligand-induced conformational change may occur and indicates potential new binding sites which are not present in any published crystal structures. The dynamic nature of the DC-SIGNR CRD may explain the wide range of ligand specificities and affinities of the C-type lectin scaffold and suggests that the study of the ligand binding properties and dynamics of proteins such as DC-SIGNR in solution is essential to further understanding of this class of proteins.

1 Introduction.

1.1 Carbohydrate Binding Proteins.

Complex carbohydrates have the potential to encode a huge amount of biological information which can be several orders of magnitude greater than that encoded in oligonucleotides and peptides. The ability to form different bond configurations (α or β) at many different positions along with variations in ring sizes produces a vast assortment of structures (Gabius et al, 2004). Glycosylated proteins in mammals are abundant in serum, the extracellular matrix and on cell surfaces and have roles in cell adhesion and molecular recognition (Drickamer, 1999; Shenoy et al, 2002). In addition, the distribution of oligosaccharides differs between mammalian and bacterial cell surfaces as well as viral envelopes and so recognition of specific glycosylation patterns aids in pathogen recognition and the discrimination between self and non-self (Larkin & Imperiali, 2011; Weis & Drickamer, 1996).

Carbohydrate binding proteins, known as lectins, have evolved to de-code the information contained within complex carbohydrate structures. In contrast to carbohydrate binding enzymes and antibodies, lectins possess no catalytic activity and are not the result of an adaptive immune response (Lis & Sharon, 1998). Lectins can vary greatly in size, structure and ligand binding profiles. However,

several sub-families exist which share similar characteristics. This project investigates a member of the sub-family of calcium-dependent (C-type) lectins.

1.1.1 C-type Lectins.

Members of the C-type lectin family contain a C-terminal carbohydrate recognition domain (CRD) which folds independently from the rest of the protein structure and is entirely responsible for the carbohydrate binding activity of the protein (Lis & Sharon, 1998). The remaining domains vary but often consist of an N-terminal transmembrane domain and 'neck' regions which project the CRD away from the cell surface. Interactions between residues in the neck domains can promote oligomerisation and the formation of higher-order clusters on cell surfaces. This clustering encourages multi-valent interactions with glycans, increasing the overall binding affinity, since monomeric CRDs tend to bind oligosaccharides with low affinity (dissociation constants are typically on the order of 2 mM for simple glycans) (Weis & Drickamer, 1996). All animal C-type lectins can be identified by a number of conserved residues in the CRD (Drickamer, 1993). In addition, the ligand binding specificity can be predicted from the CRD sequence (for example the EPN tri-peptide motif is characteristic of mannose binding, the QPD motif of galactose binding) (Drickamer, 1999). Although the number of calcium binding sites varies between C-type lectins, the primary calcium binding site is conserved (see Figure 1.2-4) (Drickamer, 1999). This calcium ion exploits the non-covalent binding ability of the monosaccharide 3- and 4-OH groups. Each OH group forms a calcium co-ordination bond in addition to hydrogen bonds with the protein side-chain (Feinberg et al, 2001; Ng

et al, 1996; Weis et al, 1992). This forms an intimate ternary complex between protein, glycan and calcium which is the hallmark of C-type lectin binding. Despite this conserved binding site and the structural similarity of C-type lectin CRDs there is huge ligand binding diversity and C-type lectins are responsible for a vast range of cellular functions (including pathogen recognition and neutralisation and receptor-mediated endocytosis) (Varki et al, 2009). The mechanisms by which the same protein scaffold can accommodate such a wide variety of interactions are not fully understood and investigations of the extended binding sites which confer specificity for particular oligosaccharides are still on-going in many cases.

1.2 Dendritic-cell Specific ICAM3 Grabbing Non-Integrin Related Protein (DC-SIGNR)

DC-SIGNR (Dendritic-cell specific ICAM3 grabbing non-integrin related) is a type II transmembrane C-type lectin that binds high mannose N-linked oligosaccharides (Feinberg et al, 2001) present on viral envelopes and host glycoproteins. In particular, DC-SIGNR interacts with the human immunodeficiency virus (HIV) glycoprotein gp120 via association with branched $Man_9GlcNAc_2$ oligosaccharides resulting in enhanced HIV infection (Bashirova et al, 2001), as well as intercellular adhesion molecule 3 (ICAM-3) (Bashirova et al, 2001) a protein involved in initiation of primary immune response, cell adhesion and cell signalling (Campanero et al, 1993; Defougerolles & Springer, 1992; Montoya et al, 2002). The interaction of DC-SIGNR with ICAM-3 along with the fact that DC-SIGNR shares 77% sequence identity with the protein DC-SIGN

(Dendritic-Cell specific ICAM3 grapping Non-integrin) (Soilleux et al, 2000), which acts as an adhesion receptor, with roles in antigen presenting, cell signalling and cell trafficking (Svajger et al, 2010), has led to speculation that DC-SIGNR may have a similar function. However, unlike DC-SIGN, which is expressed predominantly on dendritic cells (Geijtenbeek et al, 2000), DC-SIGNR is expressed on specialised endothelial cells found in the liver, lymph node, placenta and epithelial cells of the upper respiratory tract (Chen et al, 2007; Soilleux et al, 2000). As a result of the difficulties in studying these native cell types, the true physiological function of DC-SIGNR is not known with confidence but recent studies have confirmed that DC-SIGNR can act as an adhesion receptor (Yabe et al, 2010) and is involved in cell signalling which can modulate immune response (Johnson et al, 2012). There is also evidence that DC-SIGNR may act as an endocytic receptor by internalising ligands and then releasing them at low pH in the cytosol (Gramberg et al, 2008; Snyder et al, 2005b). However, DC-SIGNR contains only a single internalisation motif (while DC-SIGN contains an additional internalisation motif consisting of a tyrosine which is separated from a leucine by two amino acids; YxxL) and another report shows that DC-SIGNR/oligosaccharide binding is not pH dependent and does not internalise ligands (Guo et al, 2004).

1.2.1 Structure.

DC-SIGNR consists of four distinct domains as shown schematically in Figure 1.2-1. A short N-terminal cytoplasmic domain containing a di-leucine internalisation motif is followed by a single transmembrane domain which anchors the protein in to the cell surface. An α-helical 'neck' consisting of 7.5

highly conserved, 23 amino acid repeats projects the carbohydrate recognition domain (CRD) away from the cell surface. The neck is responsible for tetramer formation (Yu et al, 2009) while the calcium binding and carbohydrate recognition properties of DC-SIGNR are contained solely within the CRD (which exists as a monomer in solution (Mitchell et al, 2001)). As such, this project concentrates on the structure and binding properties of the CRD in isolation (the amino acid sequence of the CRD used throughout this project is shown in Figure 1.2-2).

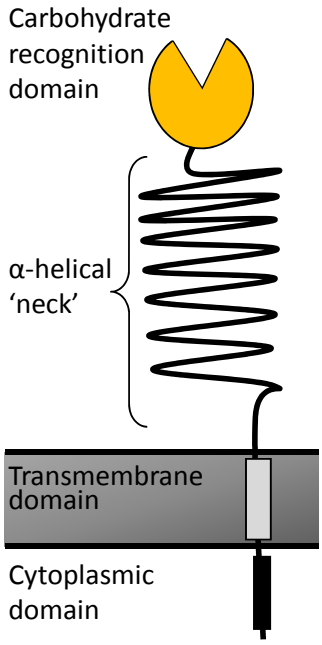

Figure 1.2-1: Schematic representation of full length DC-SIGNR. The full length protein consists of four distinct domains. A carbohydrate recognition domain is projected away from the cell surface by an α-helical neck. The neck is anchored in to the cell membrane by a single transmembrane domain followed by a short cytoplasmic domain.

```
                270                 280                 290                 300                 310                 320
AERLCRHCPK  DWTFFQGNCY  FMSNSQRNWH  DSVTACQEVR  AQLVVIKTAE  EQNFLQLQTS
          330                 340                 350                 360                 370                 380
RSNRFSWMGL  SDLNQEGTWQ  WVDGSPLSPS  FQRYWNSGEP  NNSGNEDCAE  FSGSGWNDNR
          390         399
CDVDNYWICK  KPAACFRDE
```

Figure 1.2-2: DC-SIGNR CRD sequence. The CRD sequence is numbered with the amino acid position in the full length protein. This is the fragment of the protein used throughout this project.

1.2.2 The Carbohydrate Recognition Domain (CRD).

The DC-SIGNR CRD adopts a typical 'lectin fold' consisting of α-helices and anti-parallel β-sheets connected by irregular loops which are stabilised by disulphide bonds and calcium ions (Drickamer, 1997). Thus far, X-ray crystallography has been the primary method employed for atomic-level study of the DC-SIGNR CRD structure. A number of crystal structures have been determined for the DC-SIGNR CRD: 1) in complex with the branched pentasaccharide $(GlcNAc)_2Man_3$ (PDB: 1K9J) (Feinberg et al, 2001), 2) in the absence of ligand (but in complex with one Ca^{2+}) and containing a portion of the N-terminal α-helical neck region (PDB: 1XPH) (Snyder et al, 2005a), 3) with two repeats of the α-helical neck region and one sodium ion bound (but no bound Ca^{2+}, PDB: 1XAR) (Feinberg et al, 2005); and 4) in complex with Lewis-x trisaccharide and containing a portion of the neck (PDB:1SL6) (Guo et al, 2004). There is very little variation in CRD conformation across all crystal structures (Figure 1.2-3).

Figure 1.2-3: Comparison of DC-SIGNR CRD crystal structures with and without ligand bound. The crystal structure bound to $(GlcNAc)_2Man_3$ (PDB: 1K9J) (white) and without ligand (PDB: 1XAR) (black) are shown. All published crystal structures adopt nearly identical conformations suggesting that the DC-SIGNR CRD adopts the same conformation during crystal formation with or without glycan.

Figure 1.2-4 (A) shows that the CRD binds three calcium ions, one in the primary loop (this calcium is present in all C-type lectins (Drickamer, 1999)) and two in a secondary binding loop. To date no K_D values have been reported for binding of Ca^{2+} to DC-SIGNR. However, dissociation constants for another C-type lectin, the asialoglycoprotein receptor, have been reported in the range of 330 μM (for higher-affinity Ca^{2+} binding sites 1 and 2) - 19 mM (for lower affinity site 3) (Onizuka et al, 2012). The primary calcium binding loop contains the EPN motif (a three amino acid motif consisting of a glutamic acid followed directly by a proline and asparagine) which is conserved among all mannose binding C-type lectins (Drickamer, 1992).

The primary calcium ion binds to the internal α1-3-linked mannose in $(GlcNAc)_2Man_3$ forming coordination bonds with the equatorial 3- and 4-OH groups. Amino acids that directly bind to the primary calcium also form hydrogen bonds with the ligand (Feinberg et al, 2001). This binding paradigm at the primary site is well characterised (and is the hallmark of C-type lectin binding). It has been previously indicated that the EPN sequence is responsible for mannose specificity, as mutation to QPD in the mannose-binding protein switches the specificity to galactose (Drickamer, 1992), while the location of speculated extended secondary glycan binding sites (shown in Figure 1.2-4 (B)) vary between lectins and confers specificity for larger oligosaccharide substructures (Guo et al, 2004). The extended glycan binding site of DC-SIGNR is less well characterised due to the lack of crystal structures in complex with disease associated ligands and lack of extensive mutagenesis studies. As a result, all oligosaccharides are assumed to interact with DC-SIGNR in the same way as $(GlcNAc)_2Man_3$.

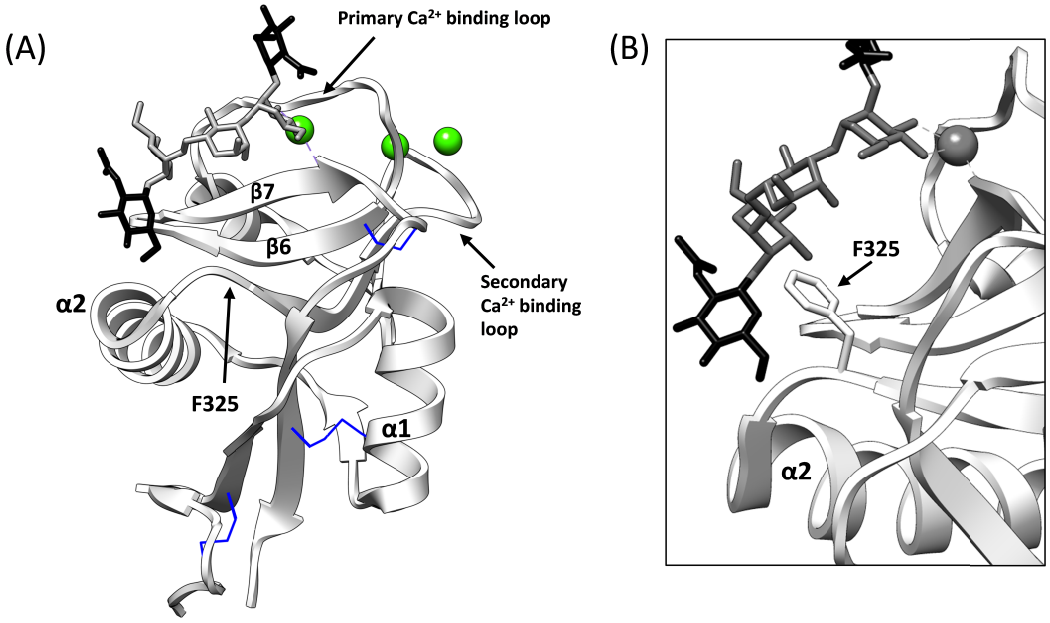

Figure 1.2-4: (A) Crystal structure of DC-SIGNR CRD in complex with $(GlcNAc)_2Man_3$ (B) $(GlcNAc)_2Man_3$ extended binding site**.** (A) The first published crystal structure of the DC-SIGNR CRD (PDB: 1K9J). Regions of interest are marked on to the CRD for reference and the three calcium ions are represented by green spheres. The mannose residues in the $(GlcNAc)_2Man_3$ glycan are light grey while the GlcNAc residues are dark grey. Disulphide bonds are highlighted in blue. (B) Highlights the extended binding site for the $(GlcNAc)_2Man_3$ glycan composed of a 'shelf' formed by α-helix 2 and the solvent exposed residue F_{325}.

The 1K9J crystal (Figure 1.2-4 (B)) shows that the extended glycan binding site for the DC-SIGNR-$(GlcNAc)_2Man_3$ complex consists of residues on α-helix 2 and a solvent exposed F_{325} residue. The C-terminal end of α-helix 2 packs against the loop joining β-sheets 6 and 7 forming a continuous binding surface. The 'shelf' formed by α-helix 2 and F_{325} creates a complementary shape to the Manα1-6Man moiety 'arm' of $(GlcNAc)_2Man_3$ which forms van der Waals contacts with F_{325} and hydrogen bonds with S_{372} (see Table 7.2-1 in Chapter 7 for a list of residues involved in binding). In addition, it is speculated that the F_{325} residue is responsible for the selective binding of DC-SIGNR to the outer branched tri-mannose moiety (grey box in Figure 1.2-5) of high-mannose structures (such as

$Man_9GlcNAc_2$) as it sterically hinders binding to the inner branched mannose (dashed box in Figure 1.2-5) (Feinberg et al, 2001).

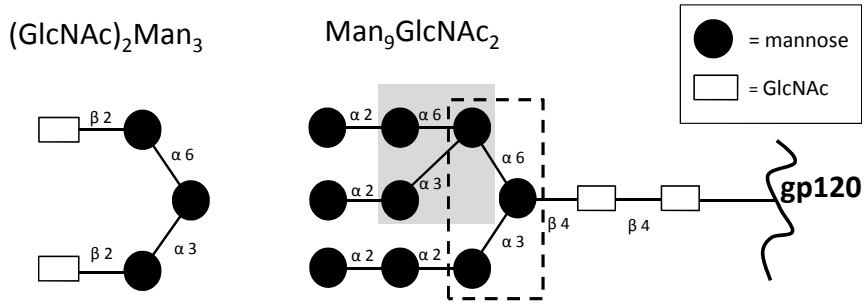

Figure 1.2-5: Schematic representation of $(GlcNAc)_2Man_3$ and $Man_9GlcNAc_2$. The $(GlcNAc)_2Man_3$ and $Man_9GlcNAc_2$ glycans are represented schematically. The inner tri-mannose structure of $Man_9GlcNAc_2$ (dashed box) is believed to be precluded from the DC-SIGNR binding site by residue F_{325} while the outer branched tri-mannose (grey box) binds to the CRD.

Although the published crystal structures provide insight in to the conformation of the DC-SIGNR CRD and the modes of glycan binding, a structure is yet to be resolved for the CRD in complex with a disease associated ligand (such as $Man_9GlcNac_2$ present on the HIV viral envelope glycoprotein gp120). Indeed, DC-SIGNR has extremely low affinity for both the ligands present in the crystal structures ($(GlcNAc)_2Man_3$ and $Lewis_X$) as neither glycan has been highlighted in glycoarray screening (Guo et al, 2004; Yabe et al, 2010)**.** As such, these glycans do not accurately represent physiologically-relevant ligands. Difficulties in obtaining crystals of DC-SIGNR CRD in complex with physiological ligands are likely due to the inherent flexibility of highly branched oligosaccharides such as Man_9 (Woods et al, 1998). Crystal structures of branched mannose glycans (Man_6, Man_4 and Man_2) in complex with the CRD of the analogous protein DC-SIGN do not contain the full ligand, suggesting that the regions of the glycan that are not in direct contact with the CRD maintain a degree of flexibility (Feinberg et al, 2007; Guo et

al, 2004). In addition, evidence of multiple binding modes suggests that the mobility/dynamics of the system may preclude formation of diffraction quality crystals.

1.2.3 Tetramer Formation.

Although individual CRDs exist as monomers and bind sugars with relatively low affinity (the K_i of mannose is ~3.2mM and $Man_9GlcNAc_2$ ~0.19mM (Mitchell et al, 2001)), the formation of the full length tetramer clusters the CRDs together increasing the affinity for binding specific patterns of oligosaccharides on the surfaces of glycoproteins (Snyder et al, 2005b). Crystal structures have been solved for the CRD containing one repeat of the neck domain (PDB: 1XPH) as well as a tetramer of neck repeats in the absence of the CRDs (PDB: 3JQH) allowing a full extracellular tetramer of DC-SIGNR to be modelled (Feinberg et al, 2009). These studies have shown that the neck forms an extended α-helical structure, that the CRD and the neck regions behave independently, and that the CRD is attached to the neck by a flexible linker which may allow the CRDs to re-orient to engage with ligands. The orientations of the CRDs may impose restraints on ligand binding, suggesting that the arrangement of the CRD cluster and disposition relative to the neck could have an effect on higher order binding specificity. Indeed, DC-SIGNR polymorphs with varying neck lengths (ranging from 3-9 repeats) have been identified in humans and hetero-tetramers of the polymorphisms have proven to be stable (Guo et al, 2006). This has led to speculation that the arrangement of CRDs with different neck lengths may affect the affinity of binding and in turn an individual's susceptibility to viral infection.

There is evidence that variations in the DC-SIGNR locus can increase HIV-RNA blood levels and sexual transmission (Xu et al, 2010), decrease viral load in hepatitis C virus (HCV) infection (Nattermann et al, 2006) and reduce rates of severe acute respiratory syndrome coronavirus (SARS-CoV) infection (Chan et al, 2006). However, the effect of DC-SIGNR polymorphism on infection susceptibility is still under debate as some authors have reported no significant association between DC-SIGNR genotype and infection (Gramberg et al, 2006; Li et al, 2008; Zhu et al, 2010).

The clustering of CRDs in the tetramer coupled with the high glycosylation of certain viral envelopes may explain the preference of DC-SIGNR to bind a range of deadly viruses including HIV (Myszka et al, 2000) and HCV (Pohlmann et al, 2003) as glycosylation patterns on pathogen surfaces are distinct from those on mammalian cells.

1.2.4 Implications in Disease.

DC-SIGNR was first investigated due to its ability to bind gp120, the glycoprotein expressed on the HIV-1 viral envelope, and increase *in trans* infection (illustrated in Figure 1.2-6) of CD4+ cells (the receptor CD4 plays an essential role in the HIV infection pathway by binding gp120 and promoting HIV cell entry (Dalgleish et al, 1984; Kwong et al, 1998)) (Baribaud et al, 2001). It has been concluded that DC-SIGNR has the potential to promote *in trans* infection of HIV *in vivo* as it is expressed on endothelial cells adjacent to CD4+ cells in the liver and placenta while CD4+ T-lymphocytes constantly migrate past the DC-SIGNR expressing sinus endothelial cells present in the lymph node (Nokta et al, 2001;

Soilleux et al, 2002). However, it is not known if HIV-1 merely exploits the antigen capturing properties of DC-SIGNR and its proximity to CD4+ cells, or if binding to DC-SIGNR triggers a conformational change in the viral envelope increasing coreceptor affinity (Myszka et al, 2000). In addition, binding to DC-SIGNR may provide a means of evading the host immune system by reducing the efficiency of antibody binding and neutralization. Expression of DC-SIGNR in the lymph node (the primary site of HIV replication *in vivo*) and placental endothelial cells further highlights the potentially devastating role of this protein in HIV infection.

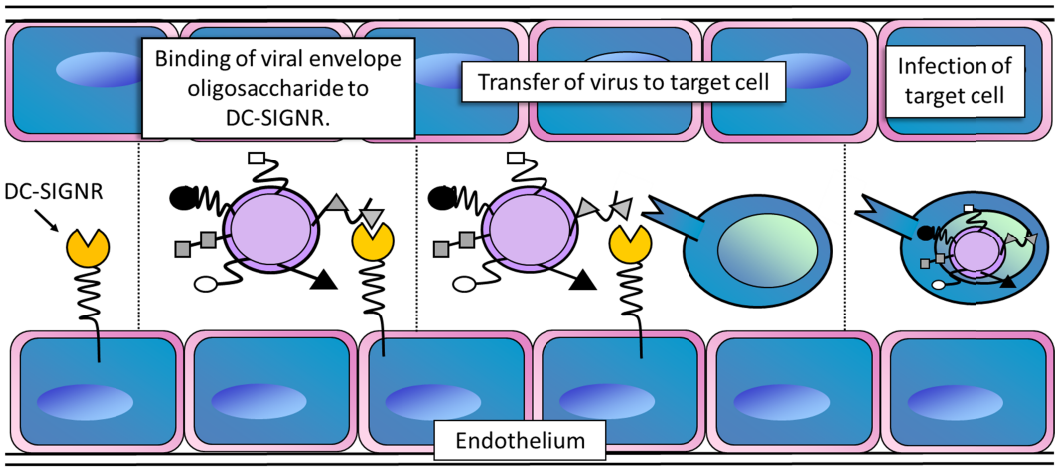

Figure 1.2-6: Illustration of *in trans* infection. DC-SIGNR enhances infection of viruses (such as HIV) *in trans*. First, DC-SIGNR binds a high-mannose oligosaccharide on a viral envelope glycoprotein. However, the virus does not infect the endothelial cell expressing DC-SIGNR and remains in an infectious state attached to the cell surface. If a 'target cell' expressing the appropriate viral co-receptor (CD4 in the case of HIV infection) is nearby, DC-SIGNR transfers the virus to the target cell which becomes fully infected. This process can enhance host infection.

It is possible that DC-SIGNR is involved in the vertical transmission (transmission of HIV from mother to child) of HIV across the placenta before birth. The mechanism of vertical transmission is not fully understood and despite interventions in the form of Caesarean sections and anti-viral therapy, vertical transmission before birth accounts for ~2% of HIV cases. Thus any protein involved in HIV binding at the maternal-foetal boundary is certainly worthy of

study (Soilleux & Coleman, 2003). DC-SIGNR is of particular interest as it has been shown to bind glycoprotein on the envelopes of several other deadly viruses including HCV (Pohlmann et al, 2003), severe acute respiratory syndrome coronovirus (SARS-CoV) (Jeffers et al, 2004) and avian H5N1 influenza (Wang et al, 2008). Also, the localisation of DC-SIGNR echoes the primary sites of replication and infection of many of the diseases with which DC-SIGNR interacts. HCV leads to chronic infection in ~70% cases and it is possible that DC-SIGNR expression in the liver sinusoidal endothelial cells contributes to this by capturing HCV and presenting the virus to adjacent hepatocytes. DC-SIGNR is also expressed in human lung type II alveolar cells, suggesting it may play a role in establishing infection at the initial sites of SARS-CoV exposure.

In addition to capturing virus and promoting *in trans* infection, DC-SIGNR is utilised by viruses such as Ebola (Simmons et al, 2003) as a direct point of entry into the host (*in cis* infection). Indeed, DC-SIGNR is exploited by a range of pathogens including viruses and mycobacteria (leprosy (Barreiro et al, 2006), tuberculosis (Tailleux et al, 2003)) in order to increase infectivity and evade the host immune system and it has also been shown to interact with parasites (Caparros et al, 2005) and tumour cells (Nonaka et al, 2008). As a result, DC-SIGNR has been classed as a universal antigen receptor.

1.3 Motivation for Research.

The ability of DC-SIGNR to interact with a range of lethal diseases and promote infection both *in cis* and *in trans* makes this a C-type lectin particularly

worthy of study. A detailed understanding of DC-SIGNR-oligosaccharide interactions could provide new therapies for a range of serious diseases for which effective treatments do not currently exist. As the association of virus to cell-surface molecules such as DC-SIGNR can be a rate limiting step in infection, these treatments would be an attractive method of decreasing the rate of infection and reducing viral load in cases where DC-SIGNR promotes infection *in trans* (HIV, HCV).

While published crystal structures provide insight into the conformation of the DC-SIGNR CRD and glycan binding mechanism, direct observation of complexes with larger, physiological oligosaccharides such as $Man_9GlcNAc_2$ remains elusive. This may be due to as yet uncharacterized conformational and dynamic factors of the CRD that prohibit crystal growth and diffraction. In addition, little is known regarding the effect of calcium binding on the CRD and, despite the extensive list of ligands, there are few direct measurements of binding affinities. In addition, the dynamics of the DC-SIGNR CRD have not been characterised. NMR has the potential to address all of these questions with the advantage that measurements can be made in solution, under near-physiological conditions.

High-field nuclear magnetic resonance (NMR) studies of DC-SIGNR CRD have not been reported, although NMR studies of ligand interactions with the homologous protein DC-SIGN have started to emerge (Angulo et al, 2008; Guzzi et al, 2011; Mari et al, 2005; Prost et al, 2012; Reina et al, 2008; Reina et al, 2007; Timpano et al, 2008). These largely ligand-based studies (i.e. properties of the

ligand are measured in free and bound states with no measurements of the changes that occur to the protein) have also been restricted to the use of small glycans and sugar mimetics and have neither approached the conformational or dynamics characterization of DC-SIGN in solution nor included larger physiological glycans such as $Man_9GlcNAc_2$. Furthermore, many of these studies typically employ molecular docking to the static crystal co-ordinates when modeling binding interactions, perpetuating the assumption that the DC-SIGN conformation in the crystals is the same for every CRD-ligand complex as well as that in solution. As a result, binding of disease-associated ligands such as $Man_9GlcNAc_2$ to molecules such as DC-SIGNR and DC-SIGN has been assumed to be analogous to the binding of smaller glycan fragments as observed in the crystalline state (Feinberg et al, 2007; Guo et al, 2004). As the lowest energy structure of smaller sugar fragments can differ from the conformation when part of a larger, branched oligosaccharide (Woods et al, 1998), the direct observation of DC-SIGNR with large oligosaccharides would be advantageous.

The goal of this project is to increase current understanding of DC-SIGNR-ligand interactions by investigating the binding of the DC-SIGNR CRD to a number of oligosaccharides in solution using heteronuclear solution-state NMR techniques that can readily deal with issues of dynamics (which may be holding back progress in crystallography). The first backbone assignment of the DC-SIGNR CRD is reported as well as the first structural data for binding of a disease-associated ligand ($Man_9GlcNAc$). These results are extended using the first measurements of transverse and longitudinal ^{15}N relaxation rates for the holo and Man_5-bound CRD. The data obtained suggest that the conformation and

dynamic properties of DC-SIGNR in solution are distinct from the existing crystal "snapshots", and that studying this protein in solution is essential for illuminating important aspects of its intrinsic physical chemistry.

1.4 Nuclear Magnetic Resonance Spectroscopy.

To date the only techniques capable of determining high resolution molecular structures are X-ray crystallography and NMR spectroscopy. NMR is particularly suitable for the study of proteins as it allows the determination of structures under close to physiological conditions in solution. In addition, NMR can be used to investigate the thermodynamic and kinetic properties of proteins, protein-ligand interactions, protein folding, and protein dynamics.

What follows is an introduction to the basic principles of NMR spectroscopy. For simplicity the more intuitive, classical representation will be used to represent the NMR phenomenon. In reality, NMR spectroscopy is based on the intrinsic, quantum mechanical property of spin which cannot be described (or visualised) by classical means.

1.4.1 Spin Angular Momentum.

All nuclei which do not contain an even number of protons and neutrons possess the quantum mechanical property spin angular momentum (or spin). Spin produces a magnetic dipole moment (μ) directly proportional to the spin angular momentum moment and so nuclei with non-zero spin can be thought of as small electromagnets. Equation (1.1) relates the nuclear magnetic dipole to the nuclear spin quantum number (I). The value of the gyromagnetic ratio (γ) is different for

different nuclear species resulting in differences in the strength of the magnetic dipole moments and differences in their NMR frequencies.

$$\mu = \gamma\hbar\sqrt{I(I+1)} \tag{1.1}$$

$$\text{where } \hbar = \frac{\text{Planck's constant}}{2\pi}$$

The application of an external magnetic field (B_0) causes the nuclear spins to orient in the direction of the external magnetic field (similar to a compass aligning with north and south). In protein NMR spectroscopy, we are primarily concerned with spin ½ ($I = \frac{1}{2}$) nuclei (such as ^{1}H, ^{13}C, and ^{15}N) because these split in to only two energy levels in the magnetic field, one parallel and the other anti-parallel to the applied magnetic field (Figure 1.4-1). The ratio of parallel to anti-parallel spins is given by a Boltzmann distribution which shows that, at thermal equilibrium, the two energy levels are nearly equally populated (as a result NMR possesses inherently low sensitivity). Nevertheless, there is a slight excess of spins parallel to the magnetic field resulting in a net macroscopic magnetization (M_0) which can be measured (Figure 1.4-1).

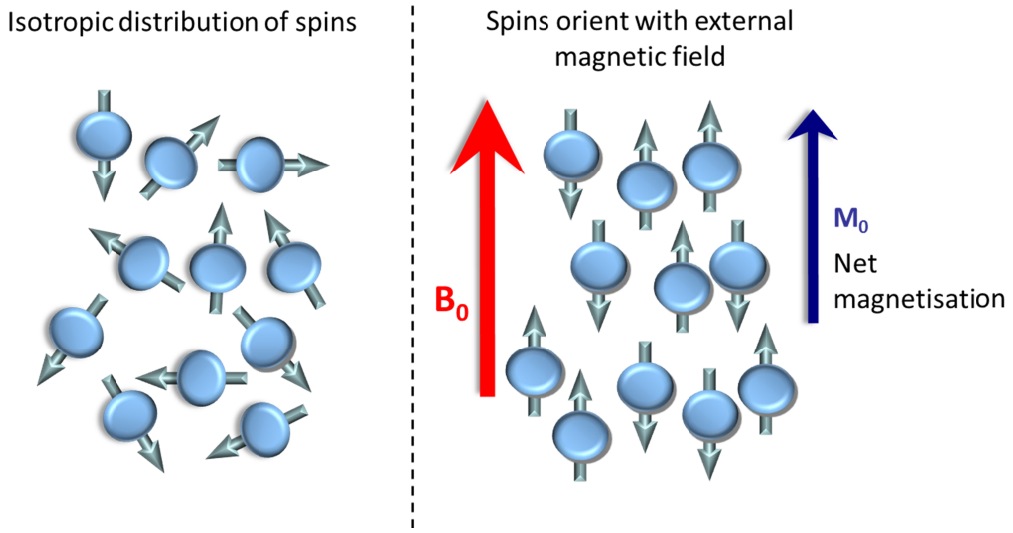

Figure 1.4-1: Distribution of spins in external magnetic field. In the absence of an external magnetic field (left) the spins are oriented isotropically. If an external magnetic field (B_0) is applied (right) spin ½ nuclei orient parallel and antiparallel to the magnetic field. There is a small excess of parallel spins resulting in a net magnetization (M_0).

1.4.2 Larmor Frequency.

If an external magnetic field is applied along the z-axis, the nuclear magnetic moments will precess about this axis as shown in Figure 1.4-2. The rate of the precession is known as the Larmor frequency (ω_0). Equation 1.2 shows that the Larmor frequency is dependent on the gyromagnetic ratio of a particular nucleus (γ) as well as the external magnetic field. Therefore, it is a physical property of the nucleus (i.e. it is different for each isotope) and dependent on the strength of the external magnetic field. It is the Larmor precession that the NMR spectrometer detects.

$$\omega_0 = -\gamma B_0 \tag{1.2}$$

However, in order to observe the Larmor precession, the magnetic moment must be perturbed away from the z-axis. This is achieved by the application of a radio frequency (RF) pulse perpendicular to the external magnetic

field. Applying an RF pulse at the Larmor frequency has the effect of rotating the equilibrium magnetization towards the xy-plane. If the RF pulse is applied for a sufficient time, the magnetization is completely transferred from the z-axis to the x-y plane (Figure 1.4-3). This pulse is called a 90° excitation pulse as the magnetization is perturbed by 90°. The magnetization is now precessing in the x-y plane which induces an electrical current that can be measured by a receiver coil.

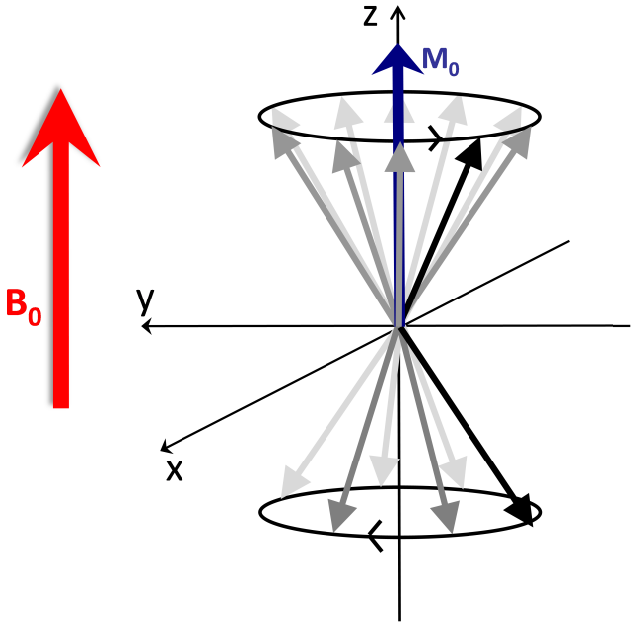

Figure 1.4-2: A Spin precessing about the magnetic field. Application of an external magnetic field (B_0) along the z-axis causes the nuclear spins to precess about that axis with net magnetization (M_0) parallel to the magnetic field.

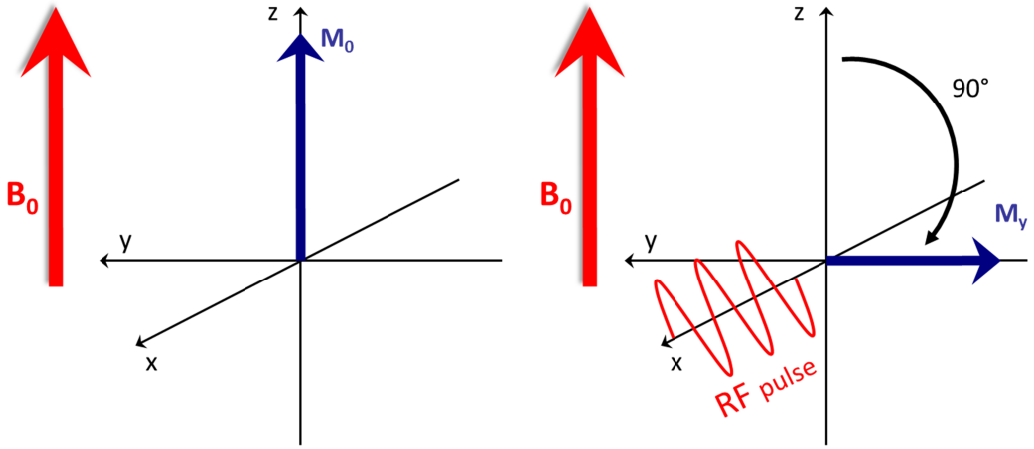

Figure 1.4-3: An RF pulse perpendicular to the external magnetic field causes the equilibrium magnetization to rotate. At thermal equilibrium the net magnetization (M_0) is along the z-axis. An RF pulse along the x-axis at the Larmor frequency causes the magnetization to rotate. If the duration of the RF pulse is sufficient the magnetization is transferred to the y-axis. The spin is now precessing in the xy-plane which induces an electrical current which can be detected.

1.4.3 Free Induction Decay (FID).

As the precession of the magnetic moment in the x-y plane represents a non-equilibrium state, the magnetization returns to the z direction (thermal equilibrium) in the absence of the RF pulse (Figure 1.4-4). This depends on the parameters R_1 and R_2 which are known as the longitudinal and transverse relaxation rates respectively. R_1 represents the return to equilibrium of the magnetization as a result of the interaction of spins with the surrounding lattice while R_2 is a result of the interactions between spins. This results in the amplitude of the detected signal decaying over time and is known as a free induction decay (FID) because it is a measure of the decay rate as a function of time (Figure 1.4-4). The FID contains the evolution of many different frequencies (different nuclei) over time. Fourier analysis converts the time dependent FID in to a function of frequency resulting in an NMR spectrum.

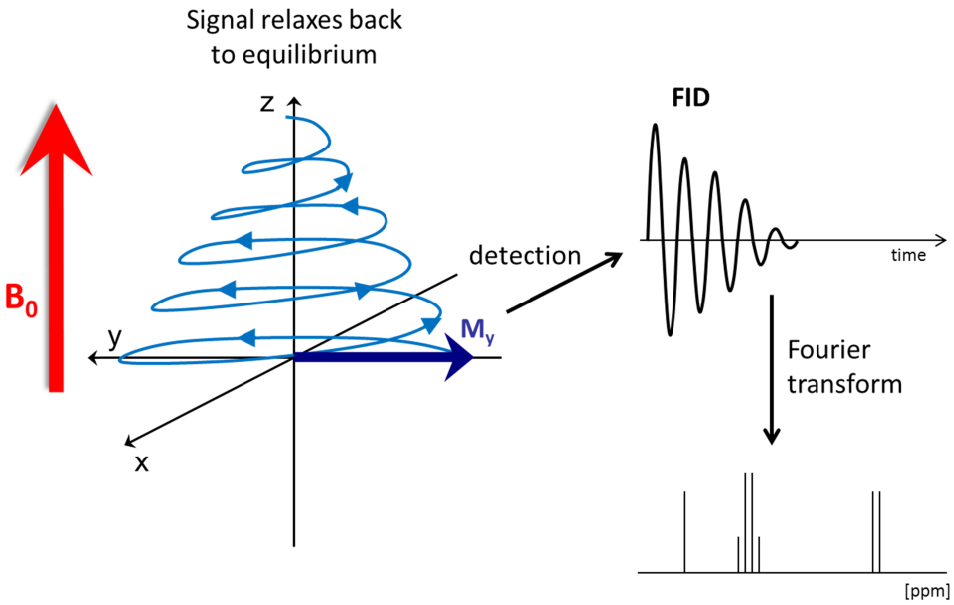

Figure 1.4-4: Free induction decay. Following the RF pulse, magnetization returns to the equilibrium state along the z-axis. The precession of the nucleus creates an electric current which is detected by a receiver coil as a free induction decay (FID). Fourier transformation of the FID transforms the time domain in the FID into a frequency domain resulting in an NMR spectrum.

1.4.4 Chemical Shift.

Because each nucleus in the magnetic field acts as a small electromagnet, each nucleus produces a local magnetic field which can oppose the external field. For example, the magnetic field produced by the electron shell surrounding a nucleus acts as a shield from the external magnetic field causing the nuclei to be at a lower resonance frequency than expected and appear shifted upfield (shifted to the right) in the NMR spectrum. In contrast, electronegative groups are de-shielded and appear downfield (shifted to the left) in the spectrum. As a result, the magnetic field experienced by each nucleus is different causing each nucleus to precess at a different frequency resulting in separation of the signals in the NMR spectrum (chemical shift). The chemical shift is an essential component of NMR spectroscopy as it contains a vast amount of information on the molecular

structure. The chemical shift allows different chemical structures (i.e. aliphatic vs aromatic protons) to be distinguished (see Figure 1.4-5) and regions of secondary structure in proteins to be identified. In addition, the distribution of chemical shifts in a protein (also called dispersion) allows the extent of protein folding to be determined. A well dispersed spectrum (many different chemical shifts) indicates that the nuclei are in many different environments, as would be expected for a protein containing elements of secondary structure (α-helix and β-sheet) along with a well-defined tertiary fold. In contrast, the nuclei in an un-folded protein all experience similar environments and so the chemical shifts are all close to their random coil values resulting in a narrow dispersion (many chemical shifts in the same region of the spectrum).

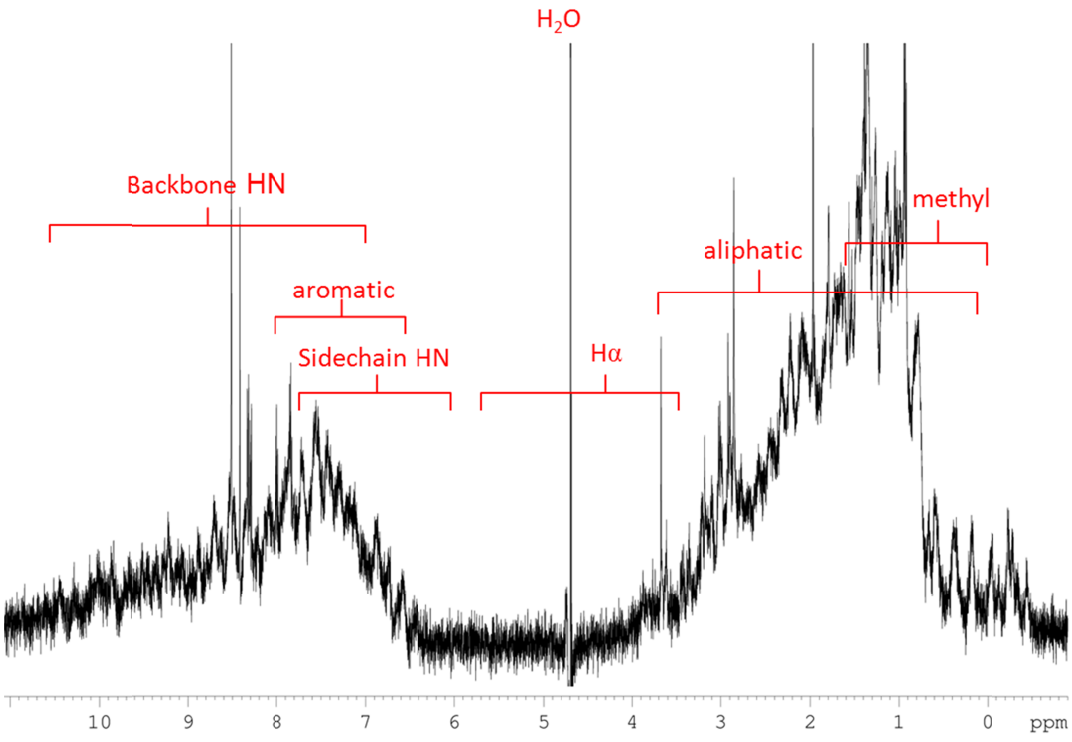

Figure 1.4-5: 1D-^{1}H spectrum of DC-SIGNR CRD. The chemical shifts of certain molecular structures are distinct and reside in defined regions of the spectrum. The chemical shifts of the CRD are well dispersed across the ^{1}H dimension suggesting that secondary structure elements are present and the protein is well folded. Due to the size of the CRD the 1D spectrum is complex, with many overlaying chemical shifts.

Figure 1.4-5 shows that, due to the size of the DC-SIGNR CRD (139 amino acids), the 1D-^{1}H spectrum contains many overlapping signals and thus is far too complex to interpret. For this reason two-dimensional (2D) and three-dimensional (3D) heteronuclear (^{15}N and ^{13}C) experiments are utilised. A 2D experiment can separate overlapping peaks in 1D spectra (for example the peaks in Figure 1.4-5) by the addition of a second dimension. Throughout this thesis we are only concerned with 2D experiments which correlate two 1D spectra (through a direct bond or through space) as shown in Figure 1.4-6. A common example of this is the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum which only contains amide signals. The HSQC experiment and 3D experiments are described in more detail below.

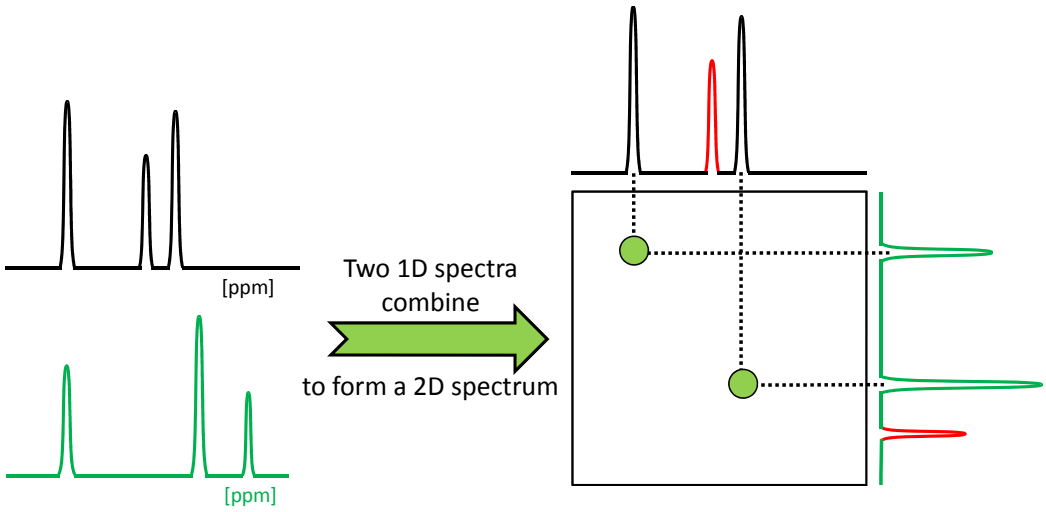

Figure 1.4-6: Simplified representation of a 2D experiment. 2D experiments can be created by combining two 1D spectra (i.e. a ^{1}H and ^{15}N spectra). In this case, peaks in the 2D spectrum are only present if there is a correlation between peaks in the 1D spectra (either through a direct bond or through space depending on the experiment used). If a peak in the 1D spectrum does not correlate with anything in the second 1D spectrum (red peaks) then there is no corresponding peak in the 2D spectrum.

1.5 NMR Assignment Strategy.

In order to study conformational and dynamics changes in the CRD of DC-SIGNR upon interaction with ligands, an assignment of the protein backbone residues is required. Due to the size of the protein, 3D heteronuclear triple resonance experiments will be used to assign sequentially the protein backbone. This section gives a brief introduction to the NMR experiments used for the amino acid assignment and the sequential assignment method.

1.5.1 The Heteronuclear Single Quantum Coherence (HSQC) Experiment.

A good quality heteronuclear single quantum coherence experiment (HSQC) (Davis et al, 1992) is a prerequisite for successful protein assignment. The ^{1}H-^{15}N HSQC spectrum of DC-SIGNR shows coupling between hydrogen and nitrogen atoms in amide functional groups found in the protein backbone and side-chains (Figure 1.5-1). This spectrum contains one signal for each amino acid due to the backbone amides, and additional signals for side-chain amides. The ^{1}H and ^{15}N chemical shifts of each peak are determined by the chemical environment of the amide group, therefore the HSQC spectrum depends upon the fold of the protein. For this reason, the HSQC spectrum is often referred to as the protein's 'fingerprint' and a well dispersed spectrum (like the spectrum of the DC-SIGNR CRD in Figure 1.5-1) is characteristic of a well-folded protein.

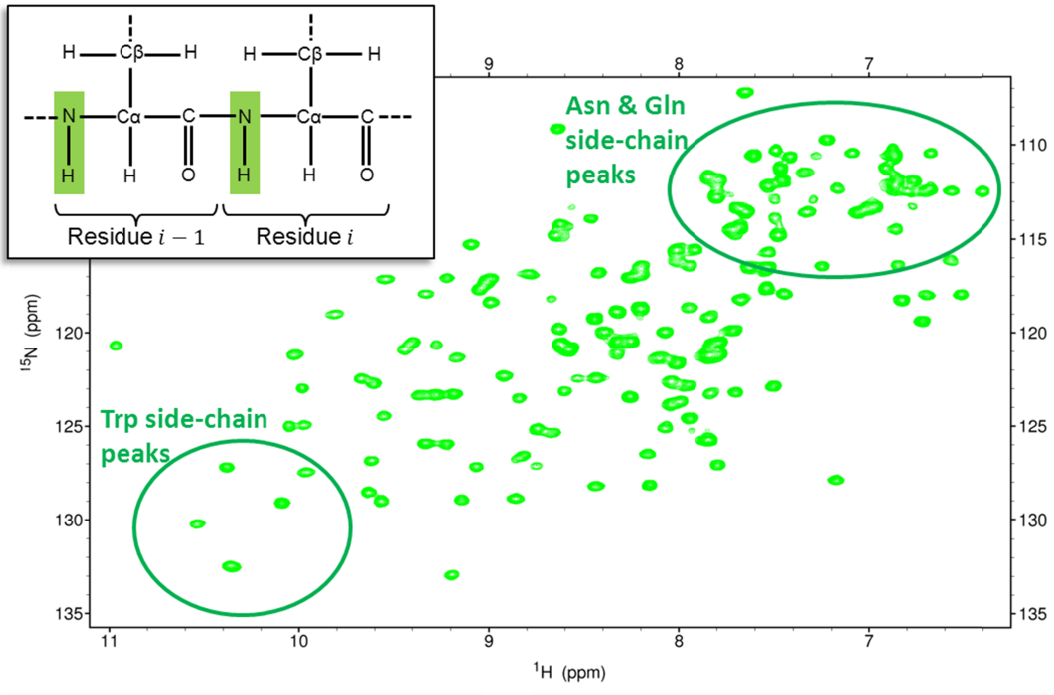

Figure 1.5-1: The heteronuclear single quantum coherence (HSQC) experiment. A spectrum of double labelled DC-SIGNR CRD at 37°C is shown as an example. Atoms in each residue which are correlated by the experiment are highlighted in green. The spectrum contains one peak for each amino acid in the protein due to the backbone amide with some extra peaks due to the side chain amides.

The HSQC spectrum can also be used as a 'map' to detect conformational changes in residues affected by ligand binding. This is an ideal method for identifying protein residues involved in DC-SIGNR/oligosaccharide interactions because, if the chemical environment of an amino acid changes (either due to ligand binding or some change in the protein conformation or dynamics as a result of binding), the position of the peak corresponding to that amino acid will change. In order to identify which amino acid has changed chemical environment upon ligand binding, each peak in the HSQC spectrum must be assigned to an amino acid in the protein sequence. As the DC-SIGNR CRD is relatively large (at 139 amino acids in length), there are several overlapping peaks in Figure 1.5-1 and it is not possible to achieve a full assignment from the HSQC data alone. For this reason, triple resonance (3D) NMR experiments are used.

1.5.2 Triple Resonance Experiments for Sequential Backbone Assignment.

Triple resonance experiments are used in order to extend the two dimensional HSQC spectrum into a third (^{13}C) dimension (illustrated in Figure 1.5-2). This allows overlapping peaks in the ^{1}H-^{15}N HSQC spectrum to be unambiguously resolved in the ^{13}C dimension, with the added advantage that side chain carbon chemical shifts give further information that can be used to identify the amino acid type. Figure 1.5-3 shows the average carbon chemical shifts (taken over all proteins deposited in the biological magnetic resonance data bank) for the C_α and C_β of each amino acid (Ulrich et al, 2008). Although many amino acids have similar carbon chemical shifts, some are easily identified and can be used to start the assignment process. For example, serine, threonine, glycine and alanine have characteristic C_β chemical shifts.

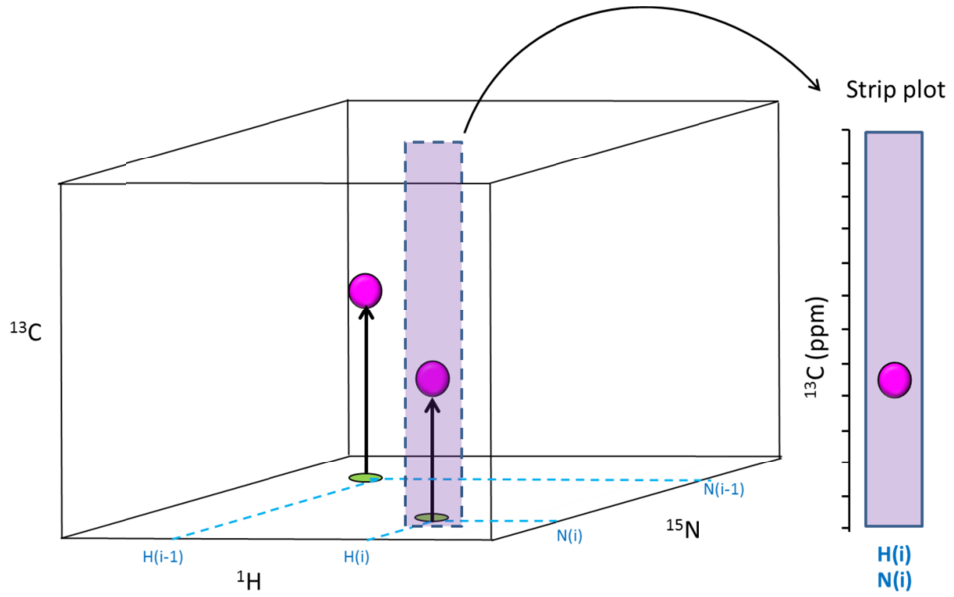

Figure 1.5-2: Simplified representation of a 3D experiment. A simplified HSQC experiment with only 2 peaks (green) is extended upwards into a third carbon dimension to form a 3D data set. The analysis of the 3D data set can be simplified by using 'strip plots': the entire carbon dimension is shown at a single ^{15}N and ^{1}H chemical shift position. In this way, only carbon signals arising from the HSQC peak at chemical shift position (H(i),N(i)) are shown.

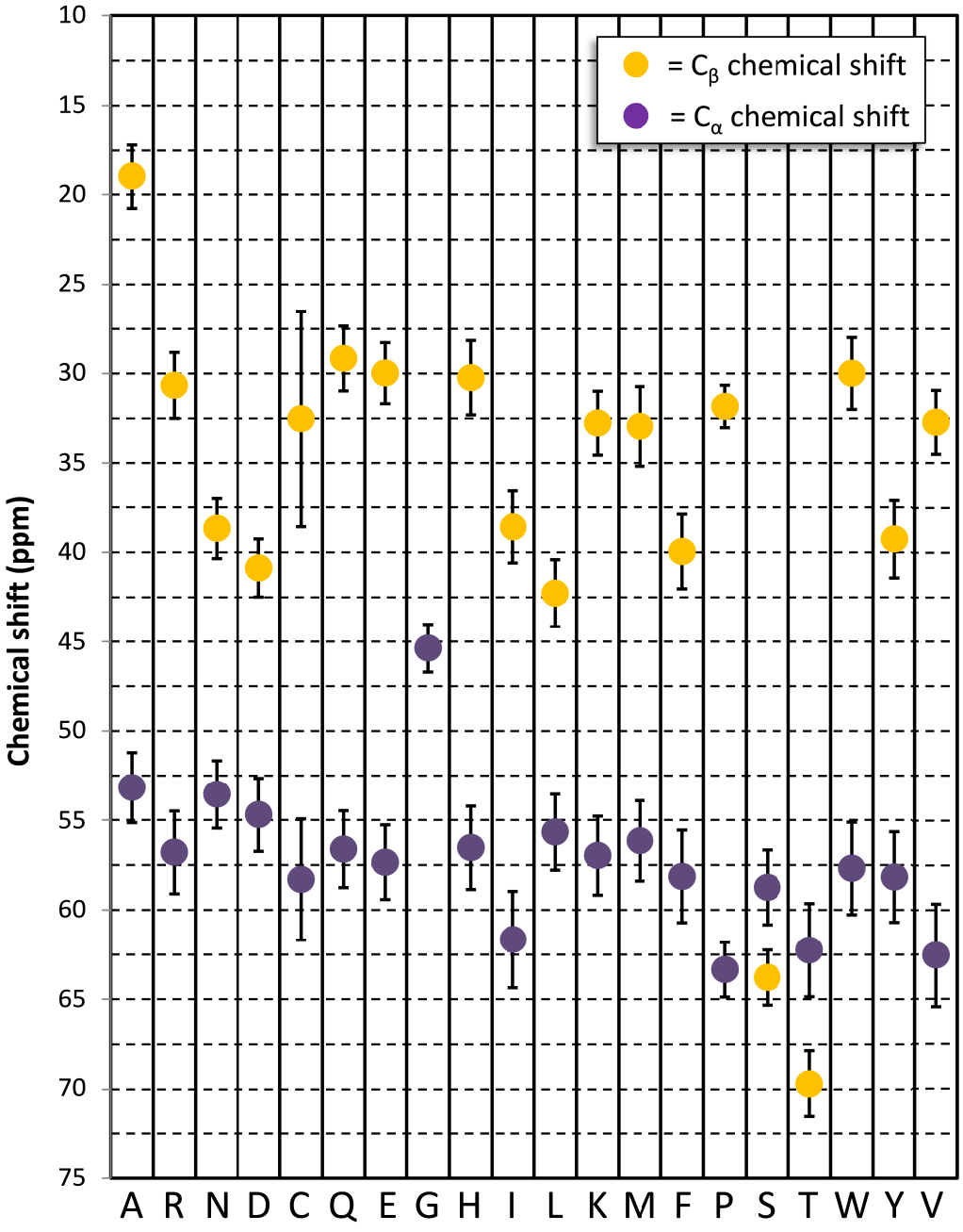

Figure 1.5-3: Average α and β carbon chemical shifts. Average chemical shift values and standard deviations (represented by error bars) are taken from the biological magnetic resonance data bank. Some amino acids are easily identified by the C_β chemical shift and can be used as a starting point for sequential assignment. Serine and threonine have high C_β chemical shift values, alanine has a low C_β chemical shift value and glycine has no β-carbon.

Triple resonance NMR experiments are named according to the atoms involved in the magnetization transfer pathway. Atoms that are involved in the

transfer pathway but not shown in the spectrum are listed in parentheses. For example, the CBCA(CO)NH (Grzesiek & Bax, 1992) experiment shows correlations between the backbone amide hydrogen and nitrogen atoms, the α-carbon and the β-carbon (Figure 1.5-4). The magnetization is transferred via the carboxyl group, but a peak for this carbon is not seen in the spectrum. For each amino acid (i), the CBCA(CO)NH spectrum correlates the backbone amide of (i) with the two carbon atoms (the C_α and C_β) from the previous amino acid ($i-1$) (as the magnetization is being transferred 'backwards' via the CO). The result is a 3D dataset with two carbon signals at the same nitrogen and hydrogen chemical shifts as each peak in the HSQC spectrum (Figure 1.5-4).

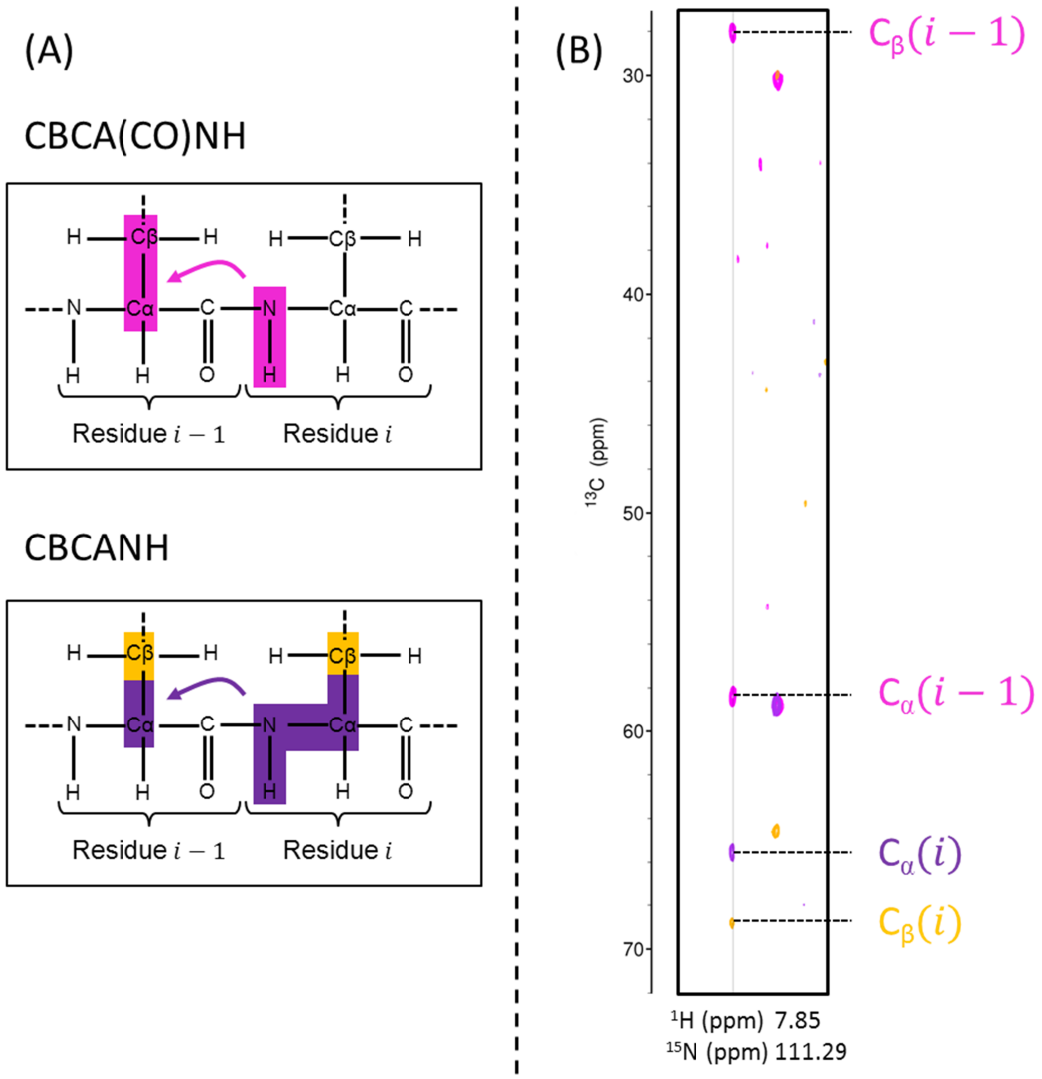

Figure 1.5-4: The CBCA(CO)NH and CBCANH experiments. (A) The magnetization transfer pathway of the CBCA(CO)NH and CBCANH experiments. During the CBCA(CO)NH experiment (pink) the amide of amino acid i is correlated with the C_α and C_β of amino acid $i-1$ while during the CBCANH experiment the amide of amino acid i is correlated with $C_\alpha(i)$, $C_\beta(i)$, $C_\alpha(i-1)$ and $C_\beta(i-1)$ resulting in four peaks. The C_β signals in the CBCANH are opposite phase allowing them to be easily identified. (B) Overlay of CBCANH and CBCA(CO)NH of double labelled, holo DC-SIGNR CRD. By overlaying the two spectra we can identify which signals belong to residue i and which to $i-1$.

In order to assign sequentially the protein backbone, triple resonance experiments are always carried out in pairs so that the carbon chemical shifts of both the i and $i-1$ residues can be unambigously identified. The CBCA(CO)NH experiment is paired with the CBCANH experiment. The CBCANH (Wittekind &

Mueller, 1993) experiment correlates the backbone amide group of residue i with the C_α and C_β atoms in the same residue as well as with the C_α and C_β of the previous residue $i-1$ (Figure 1.5-4). This results in four carbon signals being correlated to each backbone amide (four carbon signals for each peak in the HSQC). However, due to loss of magnetization, the carbon signals from residue $i-1$ are often weak/missing as seen in Figure 1.5-4 (B). Overlaying the CBCA(CO)NH spectrum with the CBCANH spectrum (Figure 1.5-4 (B)) allows the carbon signals belonging to each residue (i vs $i-1$) to be identified along with which signals arise from the C_β atoms, as these are opposite in phase and will appear negative (orange peaks in Figure 1.5-4) in the CBCANH spectrum.

Taken together, data from these experiments can be used for sequential assignment. First, a peak in the HSQC spectrum at a particular ^{15}N and ^{1}H chemical shift is selected and classified as amino acid i. Then, carbon chemical shifts for C_α and C_β atoms of amino acid i are identified using the CBCANH. If these same carbon shifts are present in the CBCA(CO)NH spectrum, the neighbouring residue in the amino acid sequence can be found (by overlaying with the CBCANH) and annotated on the HSQC as amino acid $i+1$. Using this method, chains of sequential amino acids are produced. Some of the amino acids can be identified by their carbon chemical shifts and, if a unique sequence of amino acids is found (appears only once in the protein sequence), the chain can be assigned.

Figure 1.5-5 shows strip plots of data linking a sequential chain of amino acids in the DC-SIGNR CRD via CBCA(CO)NH/CBCANH experiments. Using the

average chemical shifts given in Figure 1.5-3 as a guide, we can see that residue $i+1$ is a threonine or a serine due to the high C_β chemical shift value and that residue $i+2$ is an alanine due to its very low C_β chemical shift value. Since there are no serine residues followed by an alanine in the CRD sequence (see Figure 1.2-2), residue i must be a threonine. The amino acid sequence of the DC-SIGNR CRD in Figure 1.2-2 shows that there are only two cases of a threonine followed by an alanine and so an ambiguous assignment can be made.

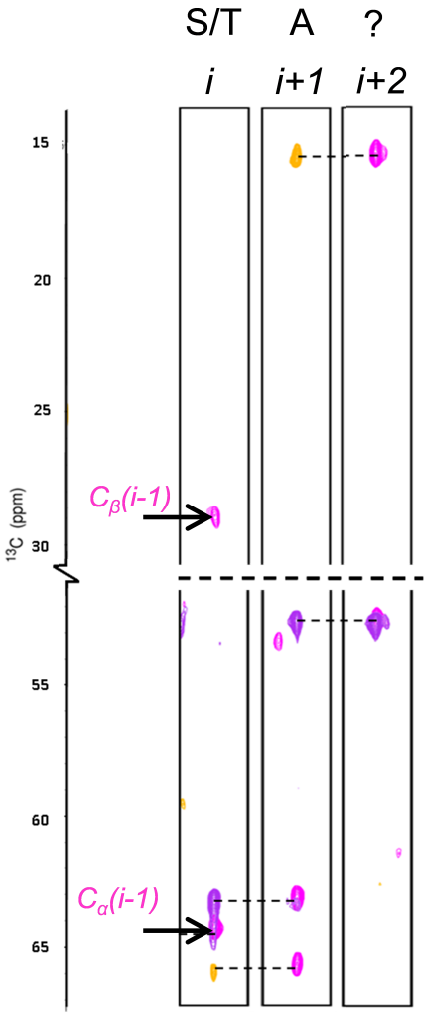

Figure 1.5-5: An example of sequential assignment using CBCANH/CBCA(CO)NH experiments**.** By overlaying the CBCANH and CBCA(CO)NH experiments, chains of sequential amino acids can be assigned. Using unique C_β chemical shifts, some amino acids can be identified. For example, the low C_β shift of strip i is characteristic of a serine or threonine while the high C_β of $i+1$ is characteristic of an alanine.

In principle, a full backbone assignment can be achieved from this pair of experiments alone. However, degeneracy (overlapping peaks) in the carbon dimension or weak signals (due to the low sensitivity of the CBCANH and CBCA(CO)NH experiments) can cause the chain to break, as seen in strip $i + 2$ of Figure 1.5-5. For these reasons, more experiments are usually required.

Figure 1.5-6 shows how the HNCA/HN(CO)CA and HNCO/HN(CA)CO experiments (Bax & Ikura, 1991; Clubb et al, 1992; Kay et al, 1990) are used to sequentially assign the protein. These experiments are more sensitive than the CBCANH as the transfer of magnetization via the one bond $^{13}C_{\alpha}$-^{15}N J coupling in the CBCANH is inefficient, particularly in proteins ~20 KDa and larger where C_{α} line widths can be greater than the ~11 Hz coupling constant (Bax, 1994). In addition the HNCA/HN(CO)CA and HNCO/HN(CA)CO experiments can be acquired with better resolution in the ^{13}C dimension in a reasonable time, given that the sweep width is smaller due to less variation in chemical shift values (as compared with the CBCANH/CBCA(CO)NH experiments) and are much more useful for forming chains of sequential amino acids. Figure 1.5-6 shows a segment of a chain of sequentially assigned amino acids in the DC-SIGNR CRD to illustrate the high signal/noise ratios of these experiments.

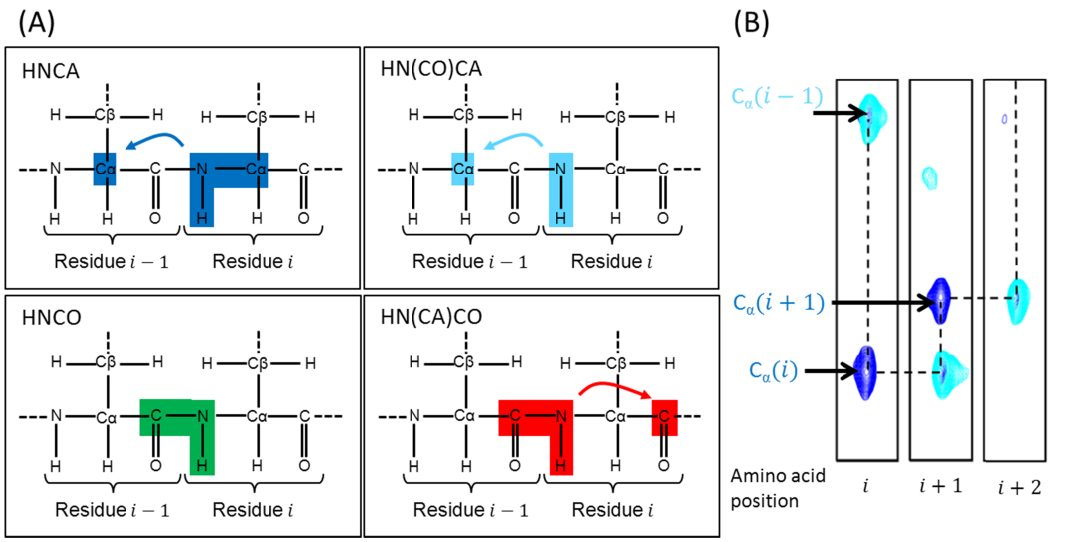

Figure 1.5-6: Triple resonance experiments for sequential assignment. (A) The magnetization transfer pathways of the HNCA/HN(CO)CA and HNCO/HN(CA)CO. For sequential assignment the experiments are run in pairs, one shows atoms in amino acid i and $i-1$ the other only atoms in amino acid $i-1$. (B) Example of sequential assignment using HNCA (dark blue) and HN(CO)CA (light blue) experiments. The method for assignment using the HNCO/HN(CA)CO experiments is the same.

However, as the C_α and CO chemical shifts are very similar for all amino acids, it is difficult to assign the spectra with these experiments alone. The C_β shift gives much more information on amino acid type, and so a combination of all six experiments is usually required for sequential assignment. In addition, as the C_α and C_β shifts are similar for the majority of amino acids, experiments that correlate all side chain carbons can be used to help identify the amino acid type once the sequential assignment has been completed. These will be discussed in more detail in Chapter 5.

1.6 Summary.

The C-type lectin DC-SIGNR has been shown to bind a range of pathogens via specific glycosylation patterns and so a detailed understanding of DC-SIGNR/glycan interactions could drive development of novel therapies in the

future. While published crystal structures of the DC-SIGNR CRD have provided valuable information of the CRD conformation and interactions with the pentasaccharide $(GlcNAc)_2Man_3$, crystals of the CRD in complex with disease associated ligands remain elusive. In addition, little is known of the effect of calcium binding or the dynamic characteristics of the CRD in solution. This project aims to address these questions using NMR spectroscopy. In order to study the CRD via NMR spectroscopy isotopically labelled $^{15}N/^{13}C$ recombinant protein has been produced in abundance. Heteronuclear experiments have been used to complete the first assignment of the DC-SIGNR CRD. This assignment facilitates the investigation of ligand and calcium binding along with the dynamic properties of the CRD and in particular the first direct investigation of a disease associated ligand ($Man_9GlcNAc$) is reported.

2 Materials & Methods

2.1 Suppliers of Reagents and Chemicals.

All reagents used in this study were of the highest grade available and were supplied by Fisher Scientific (UK) or Sigma (UK), unless otherwise stated.

2.1.1 Carbohydrates.

All carbohydrate fragments ($(GlcNAc)_2Man_3$ (M592), Man_3 (M336), Man_5 (M536), $GlcNAc_3$ (C8003)) were purchased from Dextra Labs (Dextra, Science and Technology Centre, Earley Gate, Whiteknights Road, Reading, RG6 6BZ, UK.). The $Man_9GlcNAc$ was prepared by digestion of recombinant gp120 with Endoglycosidase H using gp120 glycoprotein harvested from kifunensine-treated HEK293 cultures, followed by oligosaccharide isolation via reverse-phase chromatography and kindly supplied by Dr. Christopher Scanlan, Oxford Glycobiology Institute, Oxford, UK.

2.2 Bacterial Strains.

Table 2.2-1: *Escherichia coli* strains used in this study.

Strain	Genotype	Reference
DH5α	supE44, ΔlacU169(ϕ80lacZΔM15), hsdR17, redA1, endA1, gyrA96, this-1, relA1	(Hanahan, 1983)
BL21(DE3)	hsdS gal (λcIts857 ind1 Sam7 nin5 lacUV5-T7 gene 1)	(Studier & Moffatt, 1986)

2.3 Plasmids

Table 2.3-1: Previously generated plasmids used in this study.

Plasmid	Details	Reference
pT5T-CLEC4M	Pt5t, DC-SIGNR CRD insert sequence-Q9H2X3 (CLC4M_HUMAN), *Amp^r*	(Mitchell et al, 2001)

2.4 DNA Manipulation.

2.4.1 Preparation of Plasmid DNA.

Plasmid DNA was isolated using a QIAprep Mini-prep kit (Qiagen, Germany). The kit was used according to the manufacturer's instructions and plasmid DNA was recovered in 50 µL deionized water (dH_2O) and stored at -20°C.

2.4.2 Sequencing of Plasmid DNA.

Sequencing was performed by GATC biotech LTD, London, UK using T7 primers provided by GATC.

2.5 Growth and Maintenance of *E. coli*.

2.5.1 Rich Media

Luria-Bertani (LB) medium (10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/L sodium chloride) was used for expression of unlabelled (i.e. not isotopically-enriched) DC-SIGNR CRD. Ampicillin was added, when necessary, to a final concentration of 100 µg/mL. LB medium with 16% (w/v) agar was used for growth of *E. coli* on plates.

2.5.2 Maintenance.

For the long term storage of *E. coli* strains and constructs generated in this study, 4 mL LB cultures were grown overnight at 30°C. The overnight cultures were then used to prepare glycerol stocks by adding 1.2 mL 50% (v/v) glycerol solution. These were frozen immediately on dry ice and stored at -80°C.

2.5.3 Preparation of Super Competent E. coli Cells.

Starter cultures were prepared by inoculating 5 mL LB with *E. coli* DH5α cells and incubating overnight at 37°C with shaking at 200 rpm. The following morning, 100 mL LB was inoculated with the starter culture and cells were incubated at 25°C with shaking at 200 rpm until the absorbance at 600 nm (OD_{600}) =0.3-0.4. Cells were incubated on ice for 30 min before being pelleted by centrifugation at 1700 × *g*, resuspended in 50 ml ice cold 50 mM $CaCl_2$ and incubated on ice for 1 hour. Cells were pelleted, resuspended in 3 mL ice cold 50 mM $CaCl_2$ plus 20% (v/v) glycerol, split into 100 µL aliquots and frozen immediately on dry ice. Aliquots were stored at -80°C until needed.

2.5.4 Transformation of Super Competent E. coli Cells.

100 µL competent cells were mixed with 1-5 ng DNA in dH_2O and incubated on ice for 30 min. Cells were incubated at room temperature for 10 min before adding 500 µL pre-warmed LB and incubating at 37°C with shaking at 200 rpm for 1 hour. Cells were harvested by centrifugation at 1700 × *g*, 550 µL of supernatant was removed, the pellet was resuspended in the remaining liquid and plated on to LB-agar plates containing ampicillin (100 ug/mL). Plates were incubated overnight at 37°C.

2.6 Preparation of $^{13}C/^{15}N$ Uniformly Labelled Protein.

2.6.1 Minimal Media.

M9 minimal media was used for the expression of labelled (either singly (^{15}N) or doubly ($^{13}C/^{15}N$)) DC-SIGNR CRD. M9 media (1 L) was prepared by adding 1 g ^{15}N-ammonium chloride and 2 g ^{13}C-glucose to 200 mL of freshly autoclaved M9 salts (48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl). Next, while stirring to avoid precipitation, $CaCl_2$ and $MgSO_4$ were added to final concentrations of 100 µM and 2 mM respectively followed by addition of 10 mL 100×BME vitamin solution (Sigma, UK). The pH of the media was then adjusted to 7.3 and 100 µg/mL ampicillin added. M9 plates were prepared with addition of 16% (w/v) agar.

2.6.2 E.coli Growth in Minimal Media.

Two 100 mL starter cultures of M9 minimal media containing 100 µg/mL ampicillin and labelled reagents were inoculated with a single colony from an

M9/ampicillin plate and incubated at 37°C for 24 hours with shaking at 200 rpm. A further 100 µg/mL ampicillin was added to the starter cultures following 12 hours of growth to maintain selection pressure for the pT5T-CLEC4M plasmid. The following morning the starter cultures were used to inoculate 1 L M9 cultures (containing 100 µg/mL ampicillin) in 2 L baffled flasks (pre-warmed to 37°C) to a starting OD_{600}=0.15-0.20 and incubated at 37°C with shaking at 200 rpm.

2.6.3 Plasmid Induction.

Once an OD_{600}=0.7 was reached (typically 4 hours post-inoculation), the 1 L M9 cultures were induced with addition of isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 350 µM. Cells were induced for 20 hours at 25°C with shaking at 200 rpm before cells were harvested by centrifugation at 4000 × g for 15 minutes at 4°C. Cells were then washed with 20 ml of 10 mM Tris-HCl, pH 7.8, centrifuged at 10,000 × *g* for 10 minutes at 4°C, and the pellet frozen at -80°C until needed.

2.6.4 Protein Extraction.

Cell pellets were defrosted and resuspended in 20 ml of 10 mM Tris-HCl, pH 7.8 and 2 ml of Bugbuster™ protein extraction reagent (Merck Biosciences, UK) and lysed by sonication (30 bursts of 30 seconds using a 7 mm probe). Inclusion bodies containing protein were isolated by centrifugation at 10,000 × *g* for 15 minutes at 4°C and solubilised in 15 ml 6 M guanidine-HCl, 100 mM Tris-HCl, pH 7.8 containing 0.01% (v/v) β-mercaptoethanol by gentle rotation at 4°C for 30 mins. The sample was then centrifuged at 4300 × *g* for 30 minutes at 4°C and the supernatant (containing protein) diluted in 60 ml loading buffer (25 mM

HEPES, pH 7.8, 5 mM $CaCl_2$, 150 mM NaCl) drop by drop while stirring. The diluted mixture was then dialyzed against 2 L of loading buffer with 2 buffer changes to remove the guanidine and β-mercaptoethanol and promote protein re-folding. Following dialysis, insoluble precipitate was removed by centrifugation at 4300 × *g* for 30 minutes.

2.6.5 Expression of Selectively Un-Labelled Protein.

Selective 'un-labelling' (one amino acid in the protein lacks isotopic enrichment and so is invisible in the NMR spectra) was utilised to identify difficult to assign amino acids. A 1 L M9 culture containing 1 g/L ^{15}N-amonium chloride was supplemented with 100 mg of the amino acid of interest at OD_{600} = 0.5. The cells were then induced at OD_{600} = 0.7 as normal.

2.7 Protein Purification.

2.7.1 Mannose-Sepharose Column.

A 2 ml mannose-Sepharose column (kindly provided by Dr. Russell Wallis, University of Leicester, UK) was equilibrated with 25 mM HEPES, 5 mM $CaCl_2$, 150 mM NaCl pH 7.8 (loading buffer) before 20 mL of dialysed protein solution was loaded. Next, the column was washed with one column volume of loading buffer before eluting with 25 mM HEPES, 5 mM EDTA, 150 mM NaCl pH 7.8 in eight 1 mL fractions. Elution fractions were analysed immediately by SDS-PAGE (§2.9.1).

2.7.2 Fast protein liquid chromatography (FPLC)

Following purification by mannose-Sepharose, elution fractions containing DC-SIGNR CRD (as determined by SDS-PAGE analysis) were combined and loaded on to a Mono-Q HR 5/5 column (GE Healthcare, UK) equilibrated in 25 mM Hepes, 5 mM EDTA pH 7.8. Protein was eluted using a 500 mM NaCl gradient over 20 minutes with 1 mL fractions collected. Absorbance was monitored at 280 nm and fractions were analyzed by SDS-PAGE (§2.9.1).

2.8 NMR Sample Preparation.

2.8.1 Buffer Exchange.

Following purification, elution fractions containing DC-SIGNR CRD were exchanged into dH_2O via extensive dialysis using snakeskin™ dialysis tubing with a molecular weight cut-off of 7 KDa (Thermo scientific, UK). A volume of ~60 mL of protein was dialysed against 2 L of dH_2O with a total of 15 dH_2O changes.

2.8.2 Lyophilisation of Purified Protein.

Following extensive buffer exchange into dH_2O, purified protein was split into multiple 20 mL fractions and frozen on dry ice to form a thin layer in a 50 mL falcon tube. The tube was covered with parafilm containing holes created by a syringe needle. The 50 mL falcon tubes were briefly placed in liquid nitrogen before being attached to a vacuum. Falcon tubes were left under vacuum until all solvent was removed and the protein was stored as a dry powder at -80°C until required.

2.8.3 NMR Sample.

Lyophilised protein was dissolved in NMR buffer (20 mM deuterated Hepes (d-Hepes), 20 mM NaCl, pH 6.8 in 10% D_2O / 90% H_2O) to a final concentration of ~0.7mM. The pH was re-adjusted to 6.8 if required. Precipitate was removed by centrifugation at 1000 x g and 180 µL of sample was then transferred to 3mm Wilmad™ NMR tubes. NMR samples were stored at 4°C.

2.8.4 Determination of Protein Concentration.

Protein concentration was determined using the BCA (bicinchonic acid) protein assay kit (Thermo scientific, UK) as per the manufacturer's instructions. After incubation at 37°C for 30 min, the colour change was quantified by measuring the absorbance at 562 nm and compared to that of known protein standards (a standard curve was produced using bovine serum albumin (BSA) which was diluted in the same solvent as the DC-SIGNR samples). Figure 2.8-1 represents a typical standard curve obtained from the BCA assay. Protein concentration can be determined by calculating the gradient (m) from the standard curve in Figure 2.8-1 and using equation (2.1).

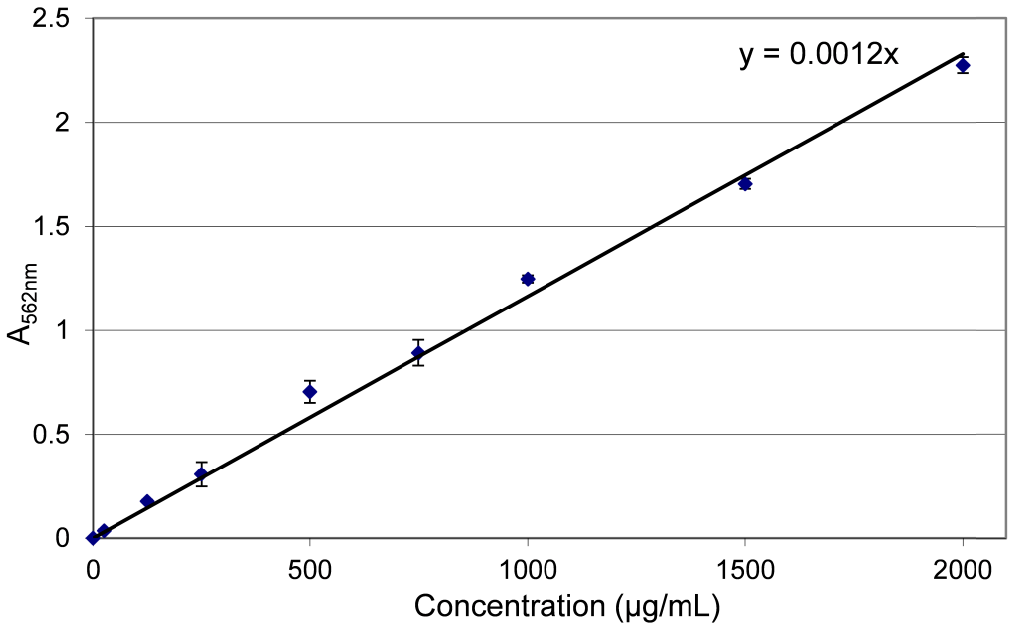

Figure 2.8-1: A typical BCA assay standard curve. A typical standard curve obtained by measuring known concentrations of BSA. The line of best fit has a gradient of 0.0012 which can be used to determine the concentration of DC-SIGNR CRD.

$$\text{Protein concentration } (x) = \frac{\text{Absorbance at 562nm } (y)}{0.0012\ (m)} \qquad (2.1)$$

2.9 Protein Resolution.

2.9.1 SDS Poly-Acrylamide Gel Electrophoresis (SDS-PAGE).

SDS polyacrylamide gels were cast and run on a vertical gel electrophoresis system according to the manufacturer's instructions, based on the method described previously (Laemmli, 1970). Typically, 0.75 mm gels were prepared with a separating gel (375 mM Tris HCl pH 8.8, 0.1% (v/v) SDS, 0.1% (v/v) APS, 0.06% (v/v) TEMED) and a stacking gel (125mM Tris HCl pH 6.8, 0.1% (v/v) SDS, 0.1% (v/v) APS, 0.06% (v/v) TEMED).

Samples were prepared by mixing with SDS sample loading buffer (125 mM Tris-HCl pH 6.8, 20% (v/v) glycerol, 4% (v/v) SDS, 0.02% (v/v) bromophenol

blue, 5% (v/v) β-mercaptoethanol) or Nu-PAGE™ sample loading buffer (Invitrogen, UK) and boiling for 5 min. Seeblue plus 2 (Invitrogen, UK) or colorplus prestained 7-175 KDa (New England Biolabs, UK) prestained molecular weight markers were used for molecular weight calibration.

Whole cell fractions taken for protein yield checks were pelleted and resuspended in 2 × SDS sample loading buffer and boiled for 10 min.

A uniform running buffer (25 mM Tris, 250 mM glycine, 0.1% (v/v) SDS) was used for electrophoresis. Gels were run at 90 mA 125 V for 15 min and for a further 35 min at 200 V.

2.10 Protein Detection.

2.10.1 Coomassie Staining.

Following electrophoresis, gels were placed in fixer (50% (v/v) methanol, 10% (v/v) acetic acid) for 30 min, then in stain solution (10% (v/v) acetic acid, 0.025% (v/v) Coomassie G250) for at least 1 hour, followed by de-stain (20% (v/v) methanol, 7% (v/v) acetic acid) until protein bands were visible.

2.10.2 Silver Staining.

Silver staining was used to screen for weak protein contaminants that are not visible using Coomassie stain. Gels were placed in fixer (50% (v/v) acetone, 1.25% (v/v) TCA, 0.015% (v/v) formaldehyde) for 15 min, followed by washing in dH_2O (three washes, then a 5 min soak, followed by a further 3 washes). Next, gels were soaked in 50% (v/v) acetone for 5 min followed by incubation in sodium

thiosulphate solution (0.017% (w/v)) for 1 min. The gel was washed in dH_2O (3 × washes) before being placed in stain solution (0.26% (w/v) silver nitrate; 0.37% (v/v) formaldehyde) for 8 min. The gel was washed in dH_2O (2 × washes) and then placed in developer (0.2 mM sodium carbonate; 0.004% (w/v) sodium thiosulphate; 0.015% (v/v) formaldehyde) until protein bands could be visualised, typically 20-30 sec. The reaction was quenched using a 1% (v/v) acetic acid solution for 1-2 min before a final wash in dH_2O.

2.11 Analysis of Purified Protein.

2.11.1 Mass Spectrometry.

Matrix assisted laser desorption/ionization (MALDI) mass spectrometry was performed on a Bruker UltraFLEX TOF/TOF instrument (Bruker, UK). Pure protein was mixed with matrix (30% (v/v) acetonitrile, 0.1% (v/v) TFA saturated with sinapinic acid) to a final concentration of ~5 µM and the instrument calibrated against lysozyme and BSA.

2.11.2 Circular Dichroism (CD).

Protein samples were prepared in 20 mM Hepes, 20 mM NaCl (the same concentrations used for NMR samples). Measurements were recorded using a J-815 spectropolarimeter (Jasco UK) using a 1 mm path-length quartz cuvette (Starna; Optiglass Ltd., Hainhault, UK). Spectra were recorded between 180 and 300 nm with a data pitch of 0.2 nm, a bandwidth of 1 nm, a scanning speed of 100 nm/min, and a response time of 1 s. The high tension of the spectra was also recorded.

2.12 Nuclear Magnetic Resonance (NMR)

All NMR experiments were carried out at 37°C (unless otherwise stated) on either a 700 MHz Bruker Avance spectrometer fitted with cryoprobe (housed at the University of Warwick) or a 5 mm triple resonance cold probe-equipped Varian Unity Inova 800 MHz spectrometer operating at a ^{1}H Larmor frequency of 799.766 MHz and a ^{15}N Larmor frequency of 81.049 MHz (HWB-NMR, University of Birmingham). Spectra recorded on Bruker instruments were referenced to residual water while spectra recorded on Varian instruments were referenced to an external DSS standard (Harris et al, 2001; Wishart et al, 1995). Bruker spectra were referenced to Varian spectra using peaks known to be stationary.

2.12.1 1D Experiments.

1D proton spectra were acquired using WATERGATE (Liu et al, 1998) or excitation sculpting (Hwang & Shaka, 1995) methods for water suppression. Spectra were acquired using 16k complex data points and 64 co-added scans. The data were zero-filled to 32k and processed using a squared sine window function.

2.12.2 2D Heteronuclear Experiments.

2D ^{1}H–^{15}N HSQC spectra (Kay et al, 1992) were acquired with 128 increments in the t_1 dimension and 1k data points in the t_2 dimension with 8 scans. The sweep width was 18.0 ppm in the ^{1}H dimension and 31.8 ppm in the ^{15}N dimension. Data were zero filled to 256 in t_1 and 2k in t_2 and processed using a squared sine window function, polynomial baseline correction in t_2 and linear prediction in t_1.

2.12.3 3D Heteronuclear Experiments.

The triple resonance (^{1}H-^{13}C-^{15}N) experiments (CBCA(CO)NH (Grzesiek & Bax, 1992), CBCANH (Wittekind & Mueller, 1993), HNCA (Kay et al, 1990), HN(CO)CA (Bax & Ikura, 1991), HNCO (Kay et al, 1990), HN(CA)CO (Clubb et al, 1992), H(CCCO)NH (Montelione et al, 1992), HCCH-TOCSY (Baldisseri et al, 1991; Bax et al, 1990), and CCH-TOCSY (Fesik et al, 1990)) were recorded with 128 increments in t_1, 40 increments in t_2 and 2k increments in the t_3 dimension. Spectra were processed using squared sine window functions and linear prediction in dimension t_1 and t_2.

3D ^{1}H-^{15}N-TOCSY-HSQC spectra (Marion et al, 1989b) were acquired with 1k, 40 and 80 increments in the t_3, t_2 and t_1 dimensions respectively with a mixing time of 60 ms.

2.12.4 Distance Constraints.

^{15}N-NOESY-HSQC (Marion et al, 1989a) were acquired with 200, 80 and 2k increments in the t_1, t_2 and t_3 dimensions respectively. Mixing times of 50, 120, 200 and 300 ms were used to plot NOE build-up curves. The 200 ms mixing time was used for structure calculation using the programme CYANA (Guntert, 2004).

^{13}C-NOESY-HSQC (Marion et al, 1989a) were recorded with mixing times of 50, 100, 120 and 150 ms.

2.12.5 Data Processing and Analysis.

Spectra were processed using Topspin 2.0 (unless otherwise stated) and analysed using CCPN Analysis software version 2.1.5 (Fogh et al, 2002; Vranken et

al, 2005) and SPARKY 3 (Kneller & Goddard, 1993). Secondary structure predictions based on the chemical shift index method (Wishart & Sykes, 1994; Wishart et al, 1992) were carried out using TALOS+ (Shen et al, 2009).

2.12.6 ^{15}N N-H Backbone Relaxation Measurements.

^{15}N longitudinal relaxation times (T_1) were measured using the inversion-recovery method as described (Farrow et al, 1994; Kay et al, 1989; Skelton et al, 1993) with relaxation delays of 0.01(×2), 0.07, 0.150(×2), 0.240, 0.340, 0.450(×2), 0.590, 0.750 seconds (×2 indicates duplicate acquisitions).

^{15}N transverse relaxation times (T_2) were measured using the spin-echo method as described (Farrow et al, 1994; Kay et al, 1989; Skelton et al, 1993) with relaxation delays of 0.0077(×2), 0.0155, 0.0232(×2), 0.0386, 0.0541, 0.0696(×2), 0.0850 seconds.

All relaxation spectra were processed in NMRpipe (Delaglio et al, 1995) and peak heights calculated and fit to a mono-exponential decay in SPARKY 3 (Kneller & Goddard, 1993).

2.12.7 Carbohydrate Titrations.

To measure binding of various carbohydrate ligands to the DC-SIGNR CRD, titration experiments were carried out by adding increasing amounts of carbohydrate (0.2, 0.5, 1.0, 2.0, 5.0, 10.0 and 20.0 mM Man_3, 1.0, 2.0, 5.0, 10.0 mM Man_5 and $(GlcNAc)_2Man_3$ and 0.1, 0.3, 0.7, 1.0, 1.5, 2.0, 5.0 and 10.0 mM $Man_9GlcNAc$) to 0.7 mM [U-^{15}N,^{13}C] - DC-SIGNR CRD at pH 6.8 and acquiring a series of 2D ^{1}H-^{15}N HSQC spectra at 37°C. The pH and temperature were held constant throughout the titration experiments. The three sugar fragments

reached saturation by 10 mM while saturation was not reached for Man_9GlcNA due to the limited amount of material available. The total chemical shift perturbation per residue ($\Delta\delta_{\text{total}}$) was calculated using the following equation

$$\Delta\delta_{\text{total}} = \sqrt{(\Delta\delta_{\text{NH}})^2 + (0.1\Delta\delta_{\text{N}})^2} \quad (2.2)$$

where $\Delta\delta_{\text{NH}}$ and $\Delta\delta_{\text{N}}$ are the chemical shift differences in the ^1H and ^{15}N dimensions respectively. The weighting factor of 0.1 applied to the nitrogen chemical shift corresponds to the difference in magnetogyric ratios of ^{15}N and ^1H nuclei. The significance of chemical shift perturbation for a given residue upon ligand addition was determined by calculating the standard deviation of the chemical shift perturbations across all residues for each carbohydrate and using this as a cut-off (Schumann et al, 2007). Perturbations that fell above this value were considered significantly different from the initial (i.e. holo) state.

2.12.8 Determination of Dissociation Constants from Titration Data.

To determine the dissociation constant (K_D), the chemical shift perturbation for well-resolved residues was plotted against ligand concentration, and the resulting curves were fitted to equation (2.3) valid for 1:1 complex formation in fast exchange regime on the NMR chemical shift time scale.

$$\Delta\delta_{(\text{total},x)} = \Delta\delta_{\max}\frac{(K_D + x + [P]) - \sqrt{(K_D + x + [P])^2 - (4[P]x)}}{2[P]} \quad (2.3)$$

where x and $[P]$ are the ligand and protein concentration respectively, $\Delta\delta_{(\text{total},x)}$ is the total chemical shift perturbation at ligand concentration x and $\Delta\delta_{\max}$ the total chemical shift perturbation upon saturation of ligand binding (Fielding,

2007; Sudmeier et al, 1980). The fit was carried out and analysed in Origin 8.5 using the non-linear least squares method.

2.12.9 $CaCl_2$, EDTA and pH Titrations.

The holo (calcium bound) DC-SIGNR CRD was investigated by addition of $CaCl_2$. Titrations of $CaCl_2$ up to 20 mM confirmed that the CRD was saturated with calcium by 4mM. In order to investigate calcium dissociation EDTA titrations (0.1, 0.3, 0.5, 1.0, 4.0, 10, 20, 40, 60 mM) were carried out on the holo DC-SIGNR CRD. For all titrations the temperature (37°C) and pH (6.8) were kept constant.

The effect of pH on calcium and glycan binding was probed by addition of 0.5 mM HCl to an NMR sample containing 4 mM $CaCl_2$ and 10 mM glycan at pH 6.8 (pH values of 7.2, 6.8, 6.0, 5.4, 4.2 and 3.2 were tested). All pH titrations were carried out at 37°C.

2.13 Specialist Software

CCPN Analysis – NMR data analysis.

CYANA – solution structure calculation.

NMRpipe – NMR data processing

Origin 8.5 – calculation of dissociation constants (non-linear least squares curve fitting)

SPARKY – NMR data analysis and assignment.

TALOS+ - secondary structure prediction from NMR chemical shifts.

Topspin 2.0 – NMR data processing.

3 Protein Preparation.

3.1 Introduction.

Although the protocol for the expression and purification of recombinant DC-SIGNR CRD in Luria-Bertani broth (LB) media has been described (Mitchell et al, 2001), the use of this protocol for expression in double labelled ($^{13}C/^{15}N$) M9 minimal media (see §2.6.1 for M9 media composition) produced extremely low yields (~350 µg/L). The triple resonance NMR experiments utilized throughout this project require protein concentrations in the range of 0.5-1 mM, which equates to approximately 3 mg of recombinant DC-SIGNR CRD per 180 µL NMR sample. Due to the high cost of isotopically labelled reagents, it is not economical nor is it practical to express multiple litres of culture for a single NMR sample. Often optimisation of the protocol will improve protein yield in M9 minimal media. This chapter describes the results of several expression trials in which the point of induction, length of induction, temperature following induction and the M9 media composition were investigated.

3.2 Optimization of Protein Yield

3.2.1 Improving Cell Growth.

The first trials in $^{13}C/^{15}N$-labelled (double labelled) minimal media exhibited extremely poor growth rates, with a 1 L culture reaching an OD_{600nm} of only 0.3 after 24 hours. While slow growth is not unusual due to the lack of nutrients in M9 when compared to rich media such as LB, it was hoped this rate of growth could be improved. Slow cell growth can be a result of improper aeration or due to the expression of protein prior to induction with IPTG, diverting cell energy from growth (this is known as 'leaky' expression). Poor aeration results in anaerobic respiration which provides 10 fold less energy than aerobic respiration and can reduce the pH of the media (which is detrimental for protein expression) (Berthold et al, 2010). To ensure sufficient aeration for aerobic respiration, 2 L baffled flasks were used for 1 L culture volumes with shaking at 200 rpm. While a 1:5 culture to flask volume ratio is recommended no substantial difference in growth rate was observed with larger flask volumes and so the 2 L baffled flasks with vigorous shaking were deemed to provide sufficient aeration. 'Leaky' expression was eliminated as a cause of poor growth because whole-cell samples taken before induction showed no CRD bands on SDS-PAGE gels (Figure 3.2-2 panel (A) lane 1). As a result, the slow growth rate was attributed to the long lag phase of BL21(DE3) cells in minimal media. Starter cultures containing 100 ml M9 grown overnight only reached an optical density of ~0.4 and so 1 L cultures were inoculated at a starting optical density <0.1. BL21(DE3) cells in M9 do not exit lag phase until an optical density of 0.1-0.2

(Berthold et al, 2010; Paliy & Gunasekera, 2007) is reached and so dilution of the overnight starter culture caused the cells to revert into lag phase resulting in the extremely slow growth rates. To overcome this, BL21(DE3) cells were grown on successive rounds of M9+ampicillin plates so that cells were conditioned to grow in minimal media. In addition, two starter cultures were grown for 24 hours at 37°C with additional ampicillin added after 12 hours to ensure selection for the DC-SIGNR plasmid was maintained. These cultures achieved an optical density of ~1.0 and were then used to inoculate a 1 L culture to a starting OD_{600nm} of ~0.2 ensuring that the 1 L culture began in exponential growth phase. These minor changes allowed the 1 L M9 cultures to reach much higher optical densities and induction at OD_{600nm} = 0.7 was achieved in as little as 4 hours following inoculation. Figure 3.2-1 shows the growth curve for a 1 L culture inoculated at OD_{600nm} = 0.15. Although these modifications substantially improved cell growth, further optimisation of the expression protocol was explored to improve the protein yield per cell. In order to improve yields, the length, temperature and point of induction were investigated along with varying M9 compositions.

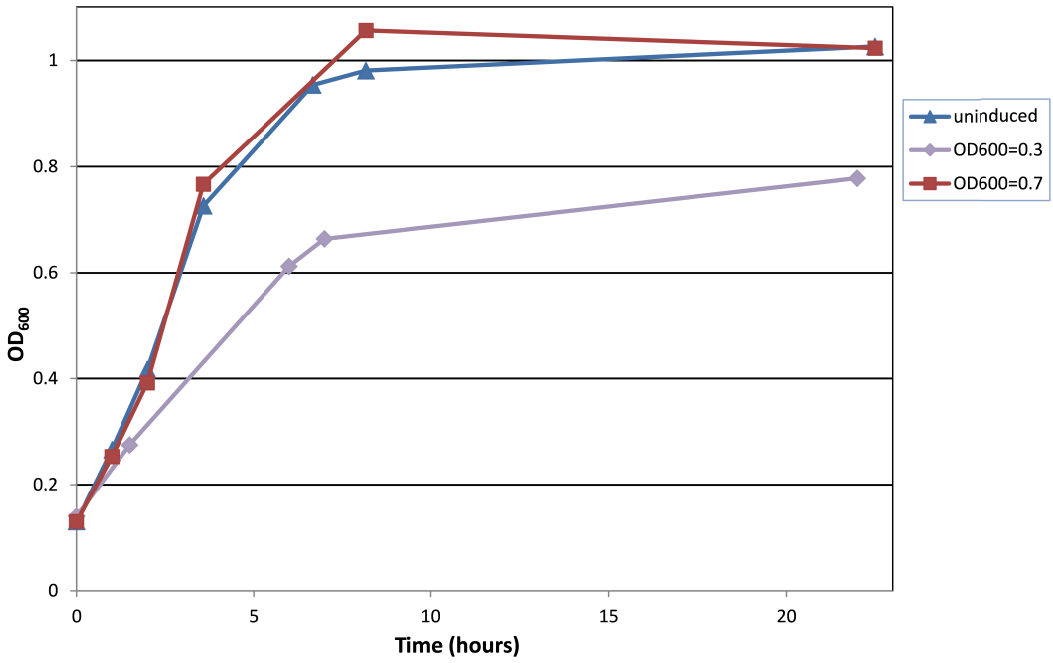

Figure 3.2-1: Growth curves of cultures induced at different optical densities. Un-induced (solid blue triangle), induced at 0.7 OD_{600nm} (solid red square) and 0.3 OD_{600nm} (solid purple diamond). Induction at 0.3 OD_{600nm} causes the rate of cell growth to immediately decline and the cells never reach the same optical densities as the un-induced or induced at 0.7 OD_{600nm} cultures. In contrast induction at 0.7 OD_{600nm} does not cause any decline in growth rate and the culture reaches the same OD_{600} as the un-induced culture.

3.2.2 Changes to Point of Induction and Length of Induction.

The standard protocol (Mitchell et al, 2001) in LB media employed an induction time of 2.5 hours, this was initially increased to 4.5 hours to compensate for the decreased growth rate caused by minimal media. However, it has been reported (Berthold et al, 2010; Paliy & Gunasekera, 2007) that substantial increases in induction time (over 40 hours) can improve yield further. The point of induction can also have an effect on the protein yield, and so we investigated inducing at the lower OD_{600nm} of 0.3 and increasing the length of induction. Figure 3.2-2 shows an SDS-PAGE gel of whole cell fractions taken at different times following induction. Each sample is normalised to the same optical density to ensure that the same number of cells are loaded in each lane.

Therefore, a more intense protein band indicates more protein is produced per cell. Both panels (A) and (B) in Figure 3.2-2 show that increasing the length of induction to 20 hours improves yield whether cells are induced at an optical density of 0.7 or 0.3. Inducing at a lower optical density does not seem to significantly affect the protein produced per cell. However, Figure 3.2-1 shows that the rate of growth declines when cultures are induced at OD_{600nm} = 0.3 (solid diamond) and these cultures never reach the optical densities of the cultures induced at OD_{600nm} = 0.7 (solid square). In contrast, cultures induced at OD_{600nm} = 0.7 show no decline in growth and reach the same optical density as un-induced cultures (solid triangle). Although, there is no substantial difference between the amount of protein produced per cell when induced at OD_{600nm} = 0.3 (Figure 3.2-2), there are fewer cells per litre (Figure 3.2-1) indicating that inducing at OD_{600nm} = 0.7 will produce higher yields. It is worth noting that the DC-SIGNR CRD migrates below the expected molecular weight of ~17 KDa. However, mass spectrometry confirms that the CRD fragment is the correct molecular weight and will be discussed later in §3.4.1.

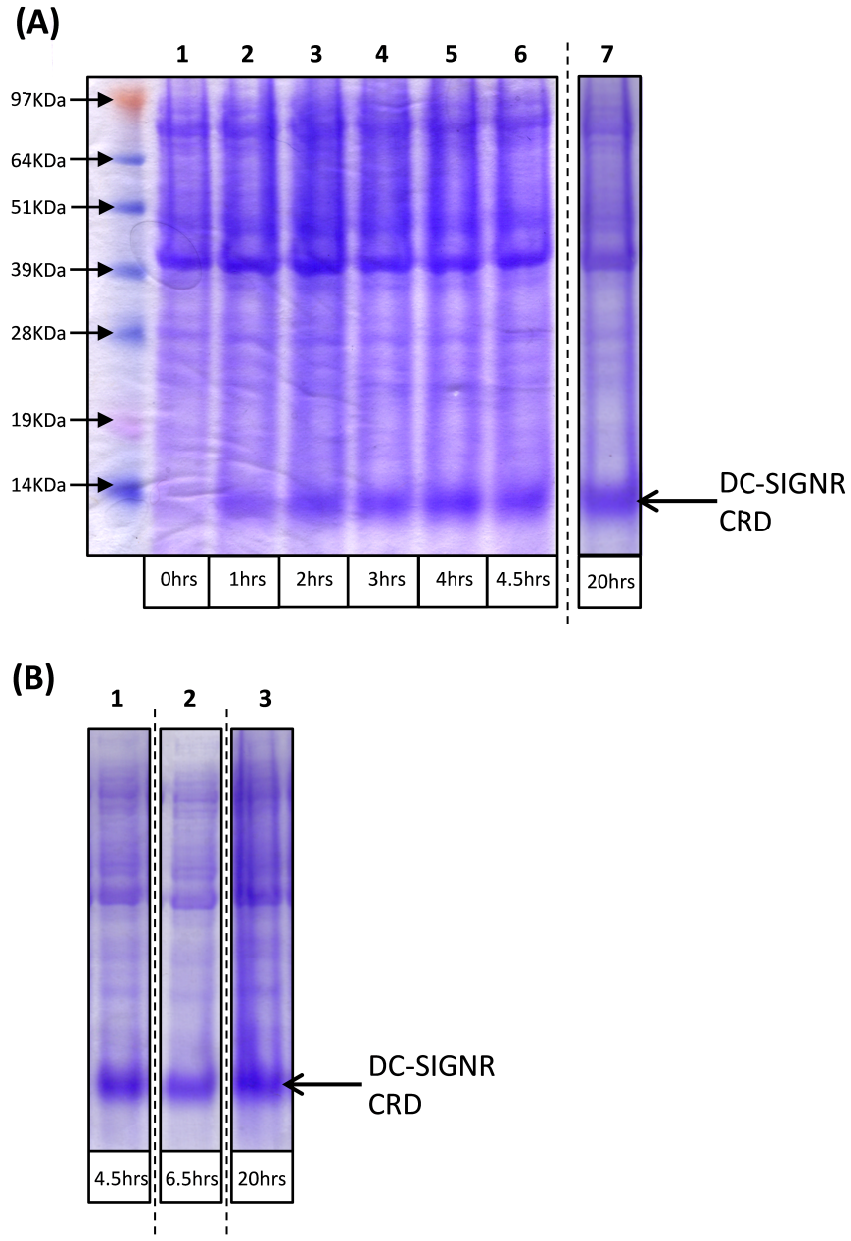

Figure 3.2-2: Investigating length and point of induction by SDS-PAGE. 1 L M9 cultures were induced with IPTG at (A) OD_{600}=0.7 and (B) 0.3 and incubated at 37°C for 20 hours. Samples were taken at varying times before (lane 1 in panel A) and after induction (time after induction is marked at the bottom of each lane). Whole cells were loaded on to SDS-PAGE gels and normalised to an OD_{600nm} = 0.4 so that the same number of cells was loaded in each lane. Dashed lines indicate that lanes are taken from separate gels/lanes were not adjacent on the gel. The band arising from the DC-SIGNR CRD is highlighted. The CRD migrates below the expected molecular weight of 17 KDa. Increasing the length of induction to 20 hours for cultures induced at both 0.7 and 0.3 OD_{600nm} improves the amount of protein produced per cell.

3.2.3 Changes in Media Composition and Temperature.

There are numerous published protocols for uniform $^{13}C/^{15}N$ labelling of protein expressed in minimal media. The addition of metal mixtures has been shown to improve yield, in particular the addition of iron alone can significantly improve expression (Paliy & Gunasekera, 2007). In addition, it has been reported that the $CaCl_2$ is not a necessary component (Berthold et al, 2010) and, as it precipitates due the presence of phosphate, its omission can improve yield. Decreasing the rate of transcription so that it is coupled with the rate of translation and post-translational events can also dramatically improve yield. This can be achieved by using weaker promoters (such as arabinose or T5) or by using *E. coli* strains containing mutations that overcome overexpression from the T7 promoter (i.e. C41(DE3) and C43(DE3)). However, the simplest method to decrease the rate of transcription is to decrease the temperature during expression. Indeed, protein production by BL21(DE3) cells can be improved by decreasing the temperature following induction to 25°C (Berthold et al, 2010). Taking the above into consideration variations in the M9 media composition along with changes in temperature were tested for improvements in cell growth and protein yield.

The standard M9 recipe (§2.6.1) was used as a control (M9), while the addition of 100 µM $FeCl_3$ ($+FeCl_3$) and the omission of calcium chloride ($-CaCl_2$) were also investigated. Two 100 mL double labelled cultures were produced using each media composition; one remained at 37°C throughout while the temperature of the other culture was reduced to 25°C following induction. Since

cultures induced at OD_{600nm} = 0.3 showed decreased optical densities (Figure 3.2-1) all cultures in this trial were induced at OD_{600nm} = 0.7. Cultures induced at 37°C were harvested after 6.5 hours (the growth curve in Figure 3.2-1 shows a slight decline in optical density at ~7 hours) while cultures induced at 25°C were left for 20 hours. Lane 1 of Figure 3.2-3 shows an un-induced (pre-ind), 100 mL double labelled culture once again confirming that there is no 'leaky' expression. Lane 2 of Figure 3.2-3 represents the control (this is the original protocol: standard M9 recipe induced for 4.5 hours at 37°C).

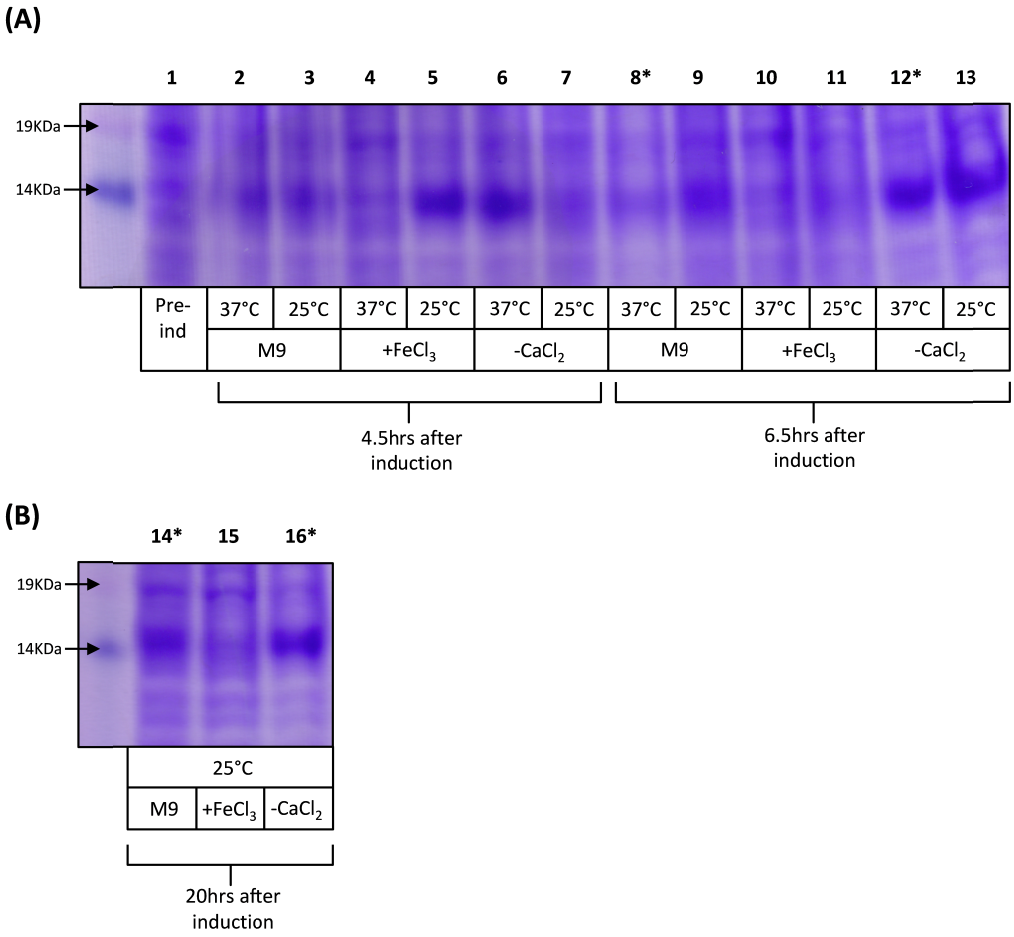

Figure 3.2-3: Investigating varying media compositions, length of induction and temperature by SDS-PAGE. Lane 1 is an example of a full cell fraction before induction to confirm absence of 'leaky' expression. The effect of addition of iron ($+FeCl_3$) and the omission of calcium chloride ($-CaCl_2$) were tested in comparison to the standard M9 recipe described in the Materials and Methods (M9). All cultures were induced at OD_{600nm} = 0.7. Two 100 mL cultures of each media composition were prepared one remained at 37°C throughout while the other was dropped to 25°C on induction. The 37°C cultures were harvested 6.5 hours post induction while the 25°C cultures were left to express for 20 hours. Samples of each culture were taken at 4.5, 6.5, and 20 hours post induction to monitor protein expression levels. Lane numbers marked with * indicate cultures which were taken forward for full purification.

First, let us consider the cultures containing the standard M9 recipe at both 37°C and 25°C (lanes 2, 3, 8, 9 and 14 of Figure 3.2-3). After 4.5 hours, there is little difference in expression between the standard M9 media at 37°C (Lane 2, Figure 3.2-3) and 25°C (Lane 3, Figure 3.2-3). However, as the length of induction increases, the 25°C culture produces more protein per cell than the 37°C culture.

The growth curve in Figure 3.2-4 also shows that the growth of the 25°C culture (solid triangle) surpasses that of the 37°C culture (solid circle). Figure 3.2-3 shows that the addition of iron improves yield at low temperatures (lanes 4 & 5) and Figure 3.2-4 shows the cultures containing iron give higher cell densities overall. However, if induced for longer than 4.5 hours, there are no DC-SIGNR protein bands visible (lanes 10, 11 and 15, Figure 3.2-3) indicating that the iron interferes with protein expression. The presence of iron may promote growth of cells without the DC-SIGNR CRD plasmid, resulting in increased growth rates of cells not expressing protein. Alternatively, iron could cause the protein to degrade possibly by preventing packing into inclusion bodies causing reduced protein expression at longer induction times. It has been shown that BL21(DE3) cells contain major defects in metal ion transport and metalloprotein biosynthesis (Pinske et al, 2011) indicating that addition of metal mixtures may have detrimental effects for this strain of *E. coli*. Removing $CaCl_2$ from the M9 recipe improves protein expression at 37°C when compared to the standard M9 recipe while omission of $CaCl_2$ at 25°C causes slower protein expression initially (lane 7, Figure 3.2-3). However, after 6.5 and 20 hours (lanes 13 and 16, Figure 3.2-3) the amount of protein per cell in the 25°C –$CaCl_2$ culture has surpassed the 37°C –$CaCl_2$ culture. This indicates that reducing the temperature after induction, omitting calcium and inducing for longer could greatly improve yield. However, the growth curve in Figure 3.2-4 shows that the optical density of the 25°C –$CaCl_2$ sample has dropped between 6.5 and 20 hours suggesting there is an optimal OD_{600nm} value between these two time points at which to harvest the cells.

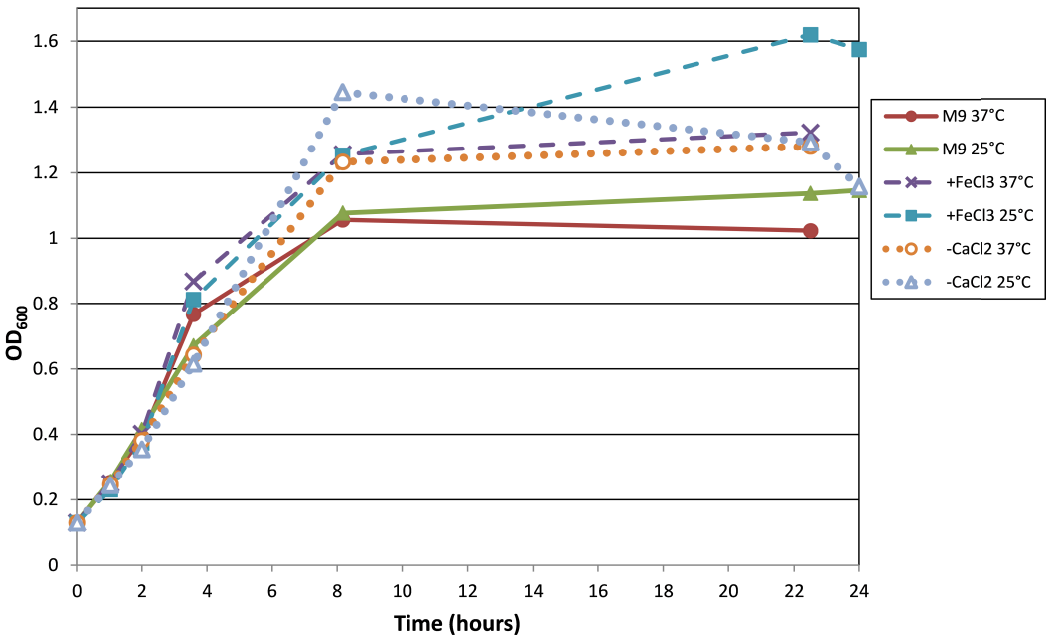

Figure 3.2-4: Growth curves of cultures with different media compositions at different temperatures post induction. All cultures show very similar growth rates prior to induction at 0.7 OD_{600nm}. Dropping the temperature after induction in standard M9 (solid green triangle) gives a small increase in the number of cells at 22 hours. The highest cell density is achieved with the addition of iron at 25°C (solid square) for 22 hours. The omission of calcium at 25°C (empty triangle) gives highest cell density at 6.5 hours after which cell density declines. Addition of iron (solid cross) and omission of calcium (empty circle) both at 37°C show similar increases in cell density at 6.5 hours which is maintained up to 20 hours.

While the use of SDS-PAGE gels is an efficient method for screening protein production per cell, it does not provide information on the fold or activity of the protein. In order to confirm that the conditions investigated above produce active protein (in the case of DC-SIGNR, protein that binds to mannose) the most promising trials (indicated by a * in Figure 3.2-3) were harvested and purified on mannose-Sepharose as described in the Materials and Methods §2.7.1. The mannose-Sepharose column is an effective method of isolating active, and therefore correctly folded, protein as it exploits the calcium dependent binding of the DC-SIGNR CRD to immobilised mannose. Protein is loaded on to the column in the presence of calcium allowing the DC-SIGNR CRD to bind to mannose while all

other proteins are eluted in the flow through and wash phases. Buffer containing EDTA was then used to elute the DC-SIGNR CRD by chelating the Ca^{2+} ions and abolishing the DC-SIGNR-mannose interaction. The yield of pure DC-SIGNR recovered from the mannose column was then quantified by BCA assay (§2.8.4) and is reported in Figure 3.2-5. Although the calcium-free cultures showed the strongest protein bands on SDS-PAGE (Figure 3.2-3) and gave higher optical densities (Figure 3.2-4), the yield of pure protein produced was lower than that of the standard M9 recipe induced for 20 hours at 25°C (Figure 3.2-5) . This suggests that although cell growth and protein expression are improved by omitting $CaCl_2$ from the M9 recipe, the protein that is produced is unable to fold correctly (during dialysis in to loading buffer as described in §2.6.4) and so cannot bind to the mannose-Sepharose column. This is supported by Figure 3.2-6 which shows the presence of DC-SIGNR protein bands in the flow through (FT) and wash (W) of the culture without calcium but not in the culture containing calcium. The presence of calcium is essential for mannose binding and is known to stabilise the protein fold so it is reasonable to conclude that the presence of calcium during expression of the DC-SIGNR CRD is required. The standard M9 recipe induced at 25°C for 20hrs gives the largest yield of active DC-SIGNR CRD (Figure 3.2-5 shows that there is almost a three fold increase compared to the original protocol) and these conditions were used in all subsequent expressions.

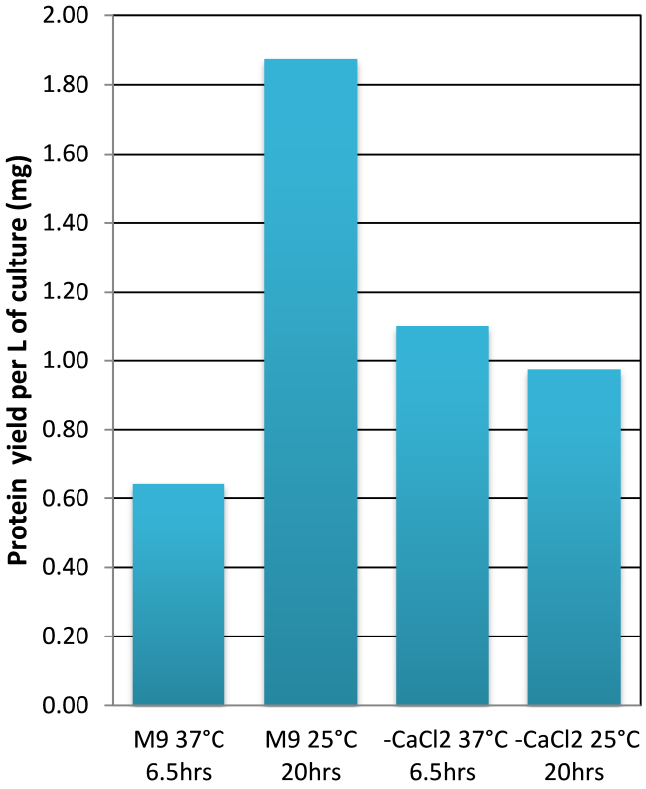

Figure 3.2-5: Yield of pure DC-SIGNR CRD per litre of culture under various expression conditions. The 100 mL cultures which showed the most promising expressions (marked with a * in Figure 3.2-3) and cell densities (Figure 3.2-4) were processed as described in the Materials and Methods and purified by mannose-Sepharose. Protein concentrations were measured by BCA assay and multiplied by 10 to give yield per L (assuming linear scaling of cell growth and protein production).

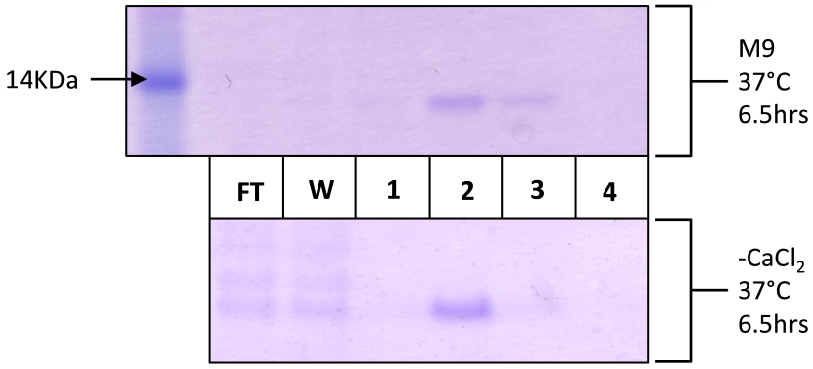

Figure 3.2-6: SDS-PAGE of mannose-Sepharose elution fractions of cultures with and without calcium. Mannose column flow through (FT), wash (W) and elution fractions (1-4) are shown from the purification of a culture of regular M9 (top) and with $CaCl_2$ omitted (bottom). Both cultures were induced at 37°C and harvested after 6.5 hours. Bands are visible in the flow through and wash of the culture without calcium suggesting the presence of mis-folded protein.

3.2.4 Improving Protein Extraction.

Evidence of DC-SIGNR was found in the insoluble pellet (P1, Figure 3.2-7) following cell lysis via sonication and solubilisation of inclusion bodies in 6 M guanidine (see Materials and Methods §2.6.4 for more detail). At this stage of the protocol, all protein should be present in the supernatant and the presence of a significant amount of DC-SIGNR remaining in the pellet (P1) suggests that whole cells or inclusion bodies remain intact. This indicated that sonication was not powerful enough to lyse the cells or that the inclusion bodies were not being completely solubilised. This was unexpected as cells grown on minimal media are typically fragile and can easily lyse (Berthold et al, 2010). Sonicating pellet P1 once more (and purifying as normal on mannose-Sepharose) extracted a significant amount of DC-SIGNR CRD (Figure 3.2-7 lane 1). However, following the second round of sonication evidence of incomplete cell lysis still remained and a strong protein band is present in pellet P2 Figure 3.2-7. This suggested that a more aggressive means of cell lysis was necessary.

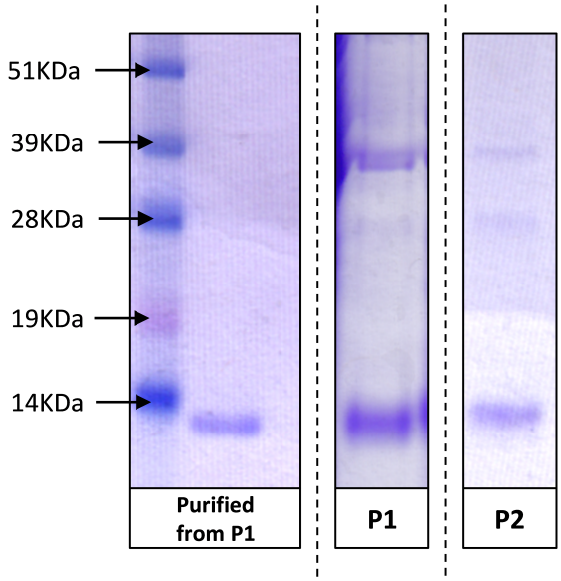

Figure 3.2-7: Protein present in pellets indicates that cells are not fully lysed. P1 is the solubilised pellet following cell lysis and inclusion body solubilisation. Protein present in pellet P1 indicates that a substantial increase in yield could be gained from complete cell lysis. Indeed, the purified lane shows protein recovered from sonicating pellet P1 for a second time and following the remainder of the purification procedure as normal. However, pellet P2 shows that there are still whole cells present following a second round of purification indicating that even more aggressive lysis would be beneficial.

The use of Bugbuster™ reagent to aid in cell lysis along with the use of a new 5 mm sonication probe and increasing the amount of sonication from 3 minutes to a total of 15 minutes resulted in almost complete cell lysis and more than doubled the total yield per 1 L culture. Further protein was recovered by limiting the amount of protein loaded on to the mannose-Sepharose column to 20 mL. Due to the higher levels of expression, the mannose-Sepharose column may become saturated resulting in correctly folded protein eluting in the wash or flow through. This was confirmed by running flow through over the column again with the result that more protein was recovered. Limiting the volume of protein loaded on to the column prevents saturation and thus all the correctly folded protein is purified in the first instance.

With the above modifications to the protocol for expression of DC-SIGNR CRD average yields of ~5 mg per L of culture were achieved. Figure 3.2-8 shows typical mannose-Sepharose elution fractions using the optimised expression and purification protocol. There is a substantial increase in the amount of protein expressed and recovered. Due to the high concentration of protein in elution fraction 2, bands corresponding to dimers are apparent. However, there is no evidence of oligomer formation under the conditions used for NMR as only one population of peaks is present in the HSQC spectrum (see Chapter 4 Figure 4.3-2) and the relaxation parameters indicate a rotational correlation time consistent with the molecular weight of the monomer (see Chapter 7 Table 7.5-1).

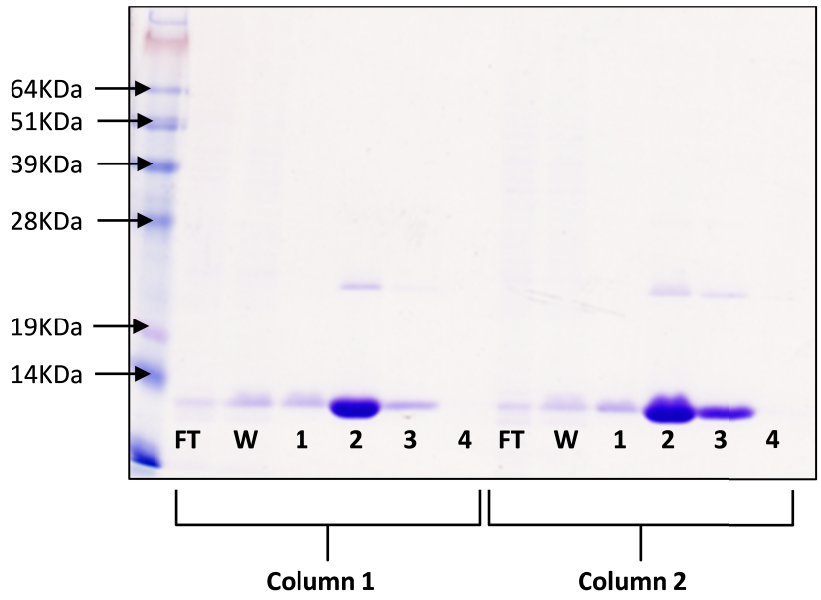

Figure 3.2-8: Elution fractions from mannose-sepharose using optimised protocol. Column flow through (FT) and wash (W) has some DC-SIGNR CRD present indicating some protein is incorrectly folded there is also evidence of other protein in these lanes which are not present in the elution fractions. Numbers correspond to 1 mL elution fractions. All protein is eluted in the first 3 mL with the most protein eluting in fraction 2. Faint dimer bands can be seen for strongest elution fractions suggesting some oligomerisation takes place at high concentrations.

3.3 Summary of Improvements to Yield.

Several modifications were made to the published protocol for expression of DC-SIGNR CRD in LB in order to optimise cell growth and expression in M9 minimal media from 350 µg to 5 mg/L culture. First, starter cultures were grown for longer and the 1 L cultures inoculated at a higher starting optical density of OD_{600nm} = 0.15 in order to compensate for the extremely long lag phase of BL21(DE3) cells in M9. This simple solution greatly improved cell growth rates allowing optical densities of OD_{600nm} = 1.0 to be reached.

Secondly, the optimal optical density at which the 1 L cultures were induced was investigated. Although there was very little difference between the amount of protein produced per cell when induced at OD_{600nm} = 0.3, or 0.7 the cultures induced at OD_{600nm} = 0.7 reached a much higher optical density overall. For this reason OD_{600nm} = 0.7 was chosen as the point of induction. Thirdly, variations of the M9 composition (addition of iron and omission of calcium) along with changes to temperature following induction and the length of induction were tested. While the addition of iron promoted higher optical densities, protein expression suffered. The omission of calcium showed greater optical densities and more protein expression per cell than the standard M9 recipe. However, the calcium free cultures produced less active DC-SIGNR suggesting that the CRD cannot correctly fold when expressed in the absence of calcium. The standard M9 recipe induced at OD_{600nm} = 0.7 for 20 hours produced the greatest amount of active DC-SIGNR. As a result, this protocol was used throughout this project. Finally, the use of Bugbuster™ cell lysis reagent, a new 5 mm sonicator probe and

increasing the amount of sonication to 15 minutes in total improved yield by ensuring extraction of all protein from cells. These modifications result in an average yield of ~5 mg of double labelled DC-SIGNR CRD per litre of culture. Yield could be improved further by investigating more points of induction (here only two were tested), the addition of more ^{13}C-glucose or ^{15}N-ammonium chloride (here the minimum amounts were used in an effort to reduce cost, but it has been shown that increasing the amount of glucose can exponentially increase protein production (Berthold et al, 2010; Marley et al, 2001)) or by growing in rich media prior to induction (although this can produce lower levels of label incorporation, this method has been successfully used for NMR analysis (Chen et al, 2006; Marley et al, 2001). However, a yield of 5 mg/L is more than sufficient for NMR analysis and so further optimisation of the protocol was not attempted.

3.4 NMR Sample Preparation.

3.4.1 Characterisation of the DC-SIGNR CRD.

The DC-SIGNR CRD migrated faster than expected on SDS-PAGE gels, so MALDI-TOF mass spectrometry was used to confirm that the protein was of the correct molecular weight and that the degree of ^{13}C/^{15}N label incorporation was sufficient for NMR analysis. The molecular weight of the unlabelled DC-SIGNR CRD obtained by mass spectrometry (Figure 3.4-1) is in close agreement with the expected molecular weight of 16194.7 Da. Figure 3.4-2 shows the results of the mass spectrometry analysis of pure, double labelled DC-SIGNR CRD. The expected molecular weight, assuming 100% label incorporation, is 17101.7 Da (907 Da

greater than the unlabelled molecular weight). The major peak in Figure 3.4-2 of 17060.695 Da indicates a label incorporation of 95.8% which is sufficient for NMR analyses. The mass spectrum confirms that the protein is of the expected molecular weight despite migrating below 14 KDa on SDS-PAGE gels. This discrepancy between the molecular weight apparent on the SDS-PAGE gel and the true molecular weight is likely due to incomplete denaturation of the protein (possibly due to the presence of disulphide bonds). The broadness of the peak indicates that there are multiple species with varying degrees of label incorporation; this is as expected for proteins expressed in labelled media. There is an additional peak 204-215 Da greater than the primary peak in both spectra which could indicate a C-terminal tryptophan modification (possibly due to a stop codon mutation UGA to UGG). If this is the case this should pose no problems to the NMR analysis as the C-terminal residues are not observable due to the high degree of flexibility. In addition, the +204 Da peak is of low intensity and a single population of peaks is present in the HSQC experiment indicating that this sample is sufficient for NMR analysis.

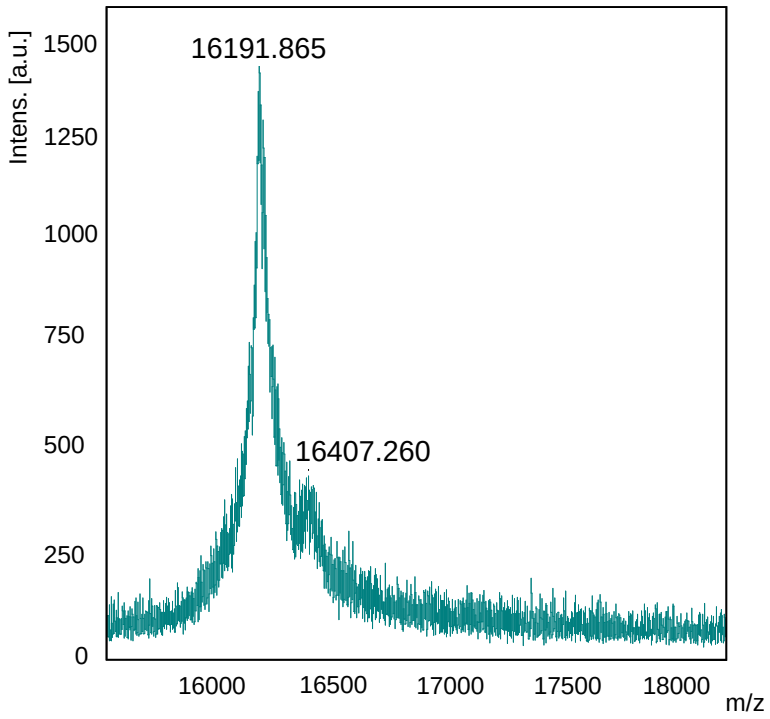

Figure 3.4-1: MALDI-TOF mass spectrometry of unlabelled DC-SIGNR CRD. The major peak of 16191.865 Da is in close agreement with the expected mass of 16194.7 Da. There is an additional, low intensity peak +215 Da from the major peak.

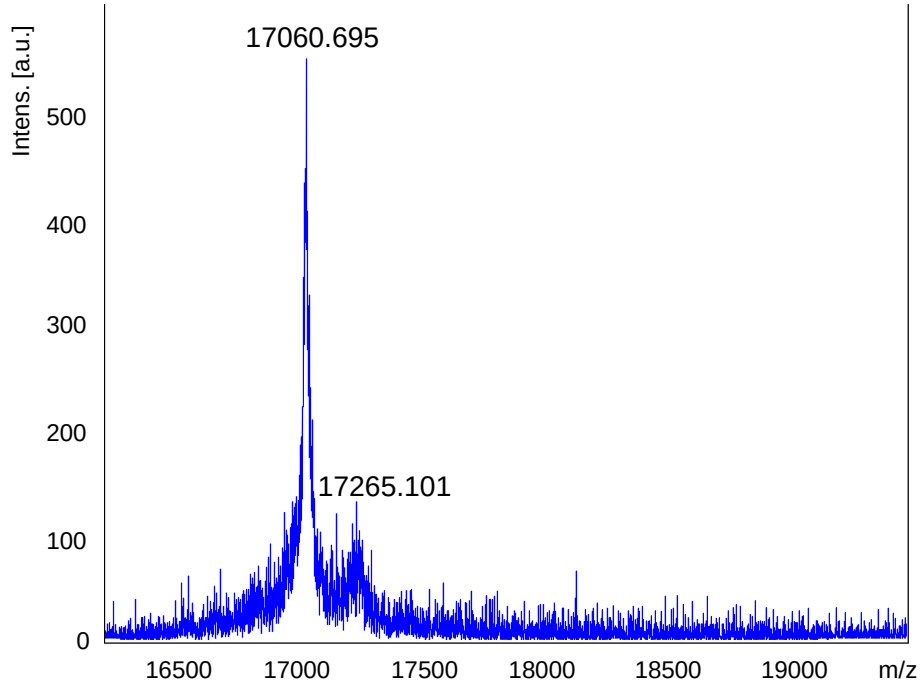

Figure 3.4-2: MALDI-TOF mass spectrometry of double labelled DC-SIGNR CRD. The major peak of 17060.695 Da is in close agreement with the expected molecular weight of double labelled DC-SIGNR CRD of 17101.7 Da. This implies an overall label incorporation of 95.5%. The broadness of the peak is likely due to varying degrees of label incorporation. There is a secondary peak +204Da from the major peak. However, peaks from this secondary mass do not appear in the NMR spectra.

CD spectra were measured of the CRD in NMR buffer (20 mM Hepes, 20 mM NaCl) in order to confirm the secondary structure of the protein. However, the presence of almost equal amounts of α-helix, β-sheet and random coil in the CRD produces a relatively featureless CD signal with weak intensity (Mitchell et al, 2001). As a result, NMR is much better suited to gauge the secondary structure content of the CRD. The HSQC spectrum (§4.3, Figure 4.3-2) is used as evidence that the protein is well folded (as dispersed chemical shifts are indicative of secondary structure elements along with tertiary fold) and secondary structure prediction using the chemical shift assignment and the chemical shift index (CSI) method (discussed in detail in §5.2) are in good agreement with the crystal structure. In addition, the protein binds to the mannose-Sepharose column confirming that the protein is correctly folded and active.

3.4.2 Determination of Purity for NMR.

Following optimisation of the expression protocol and characterisation of the purified protein, it was then necessary to ensure that the sample quality was sufficient for NMR analysis. One dimensional ^{1}H NMR spectra (Figure 3.4-3) highlighted the presence of contaminants which appeared as sharp, resolved peaks while protein signals are broader. These contaminants could be a result of the purification procedure or improper buffer exchange into NMR 'invisible' buffers (in this case deuterated Hepes (d-Hepes)). The original protocol utilised FPLC anion exchange chromatography as a second purification step following mannose-Sepharose purification. However, Figure 3.4-3 (panels (A) and (B)) shows that many contaminants were present following anion exchange. The silver

stained gel in Figure 3.4-4 (silver stain is more sensitive than Coomassie stain and so should highlight any protein contaminants) along with the mass spectrum and HSQC spectrum show that no protein contaminants are present following mannose-Sepharose purification. As such there is no need for the anion exchange purification step and so this step was omitted. Initially Vivaspin™ centrifugal concentrators were used to buffer exchange the protein in to d-Hepes. However, Figure 3.4-3 (A) shows that buffer exchange was not complete (residual un-deuterated hepes peaks are present) and that glycerol was introduced using this method. As an alternative, extensive dialysis into water was used, followed by lyophilisation and transfer in to NMR buffer. Figure 3.4-3 (C) confirms that mannose-Sepharose purification coupled with dialysis eliminates contaminants and produces good quality NMR samples. Despite extensive dialysis (against as much as 30 L of water in total) the EDTA (remaining from the mannose-Sepharose purification procedure) was not removed completely. This is consistent with previous reports which state that EDTA can be extremely difficult to remove (Pereira-Mouries et al, 2002; Searcy & Greif, 1976). While the presence of EDTA precludes the investigation of accurate calcium binding affinities, it does not interfere with the protein assignment as it is not visible in the ^{15}N edited HSQC. In addition, EDTA is useful for probing the presence of calcium in the NMR samples as the presence of two Ca^{2+} free EDTA peaks in Figure 3.4-3 (C) confirm that there is no calcium present in the sample and hence that the protein is in the apo (calcium free) form. This will be discussed further in Chapter 6.

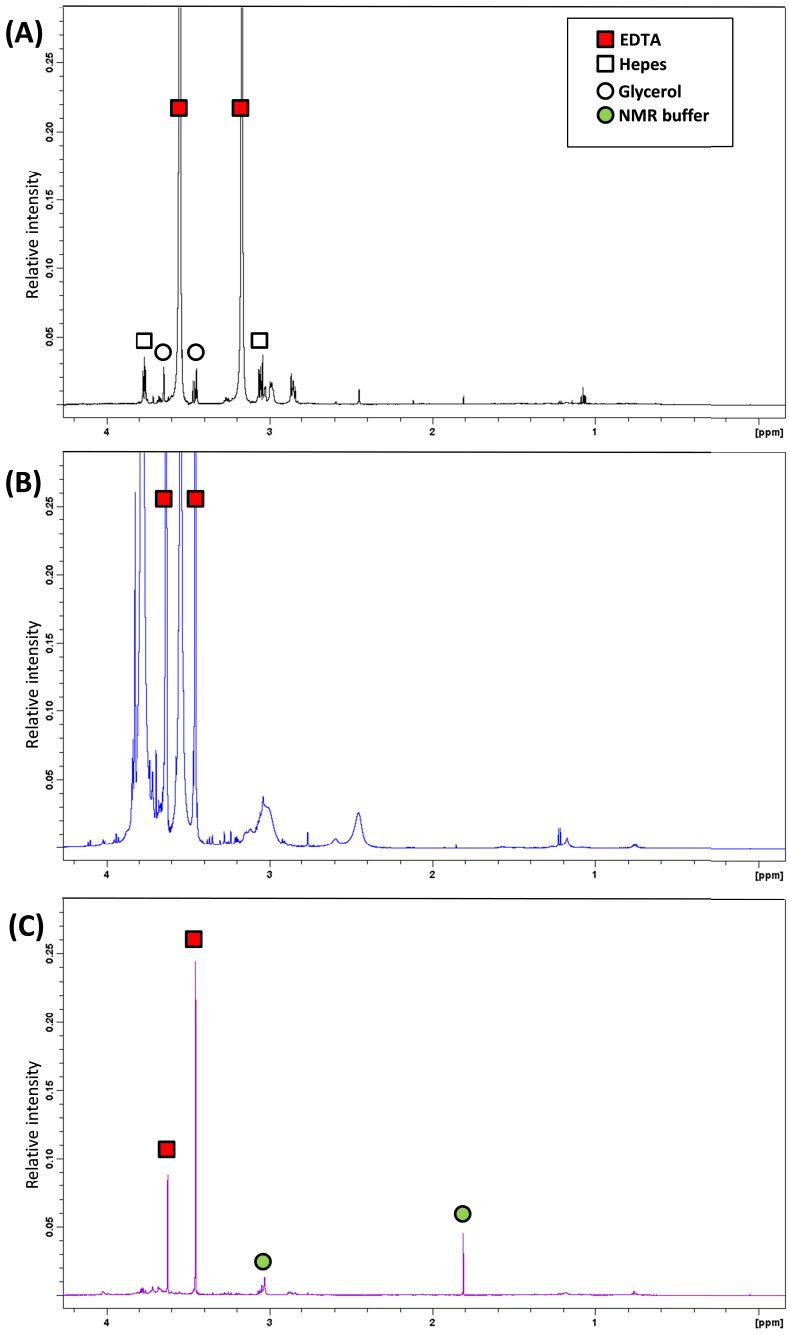

Figure 3.4-3: 1D ^{1}H NMR spectra of contaminants resulting from various purification and buffer exchange methods. Initially FPLC anion exchange chromatography following purification on mannose-Sepharose was used samples were then exchanged in to NMR buffer using (A) vivaspin™ centricons or (B) dialysis in to water followed by freeze drying. Many contaminants were present following anion exchange while centricons did not sufficiently buffer exchange as there is still hepes and a great deal of EDTA in the sample. (C) Omitting the anion exchange step followed by extensive dialysis produced the least contaminants. However, EDTA is still present.

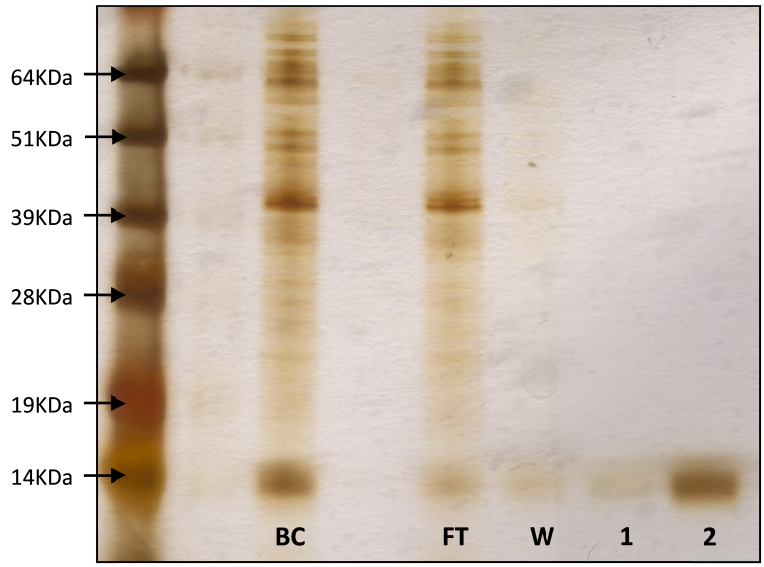

Figure 3.4-4: Silver stained gel of mannose-Sepharose column elution fractions. A sample of un-purified protein directly before loading on to the mannose-Sepharose column (BC) shows a significant amount of DC-SIGNR CRD has been expressed along with many protein contaminants. The column flow through (FT) confirms that the protein contaminants do not bind to the mannose-Sepharose. There is also some DC-SIGNR CRD in flow through which may be incorrectly folded protein. Some contaminants are also present in the wash (W). A small amount of DC-SIGNR is present in the first elution while the majority of protein is eluted in fraction 2. The lack of contaminant bands in the elution fraction when silver stained confirms the purity of the protein following mannose-Sepharose.

3.5 Conclusion.

Protein expression protocols optimised in rich media such as LB cannot always be used for expression in minimal media. The previously described protocol for expression of DC-SIGNR CRD gave very poor yields in M9 and so many changes to the protocol were investigated. Due to the extensive lag phase of BL21(DE3) cells in M9, cultures were inoculated at a higher optical density (OD_{600nm} = 0.15). While cells grown on LB can be harvested 2.5 hours following induction, the M9 cultures were left for a total of 20 hours. Allowing the cultures to induce for 20 hours along with dropping the temperature following induction to 25°C provided the greatest increase in yield of active DC-SIGNR CRD. The use of Bugbuster™ cell lysis reagent along with modifications to the sonication step

increased yield further by ensuring that all protein is extracted from the cells. In total the modifications described in this chapter produce, on average, 5 mg/L of active DC-SIGNR CRD.

Following optimisation of the expression protocol the purified DC-SIGNR CRD was characterised via mass spectrometry. This confirmed the protein was at the expected molecular weight (despite running at a lower molecular weight on SDS-PAGE gels) and that 95% of the $^{13}C/^{15}N$ labels had been successfully incorporated in to the protein. Initial NMR sample preparations utilising anion exchange showed evidence of many contaminant peaks in the ^{1}H NMR spectrum. As silver staining and mass spectrometry confirmed that no protein contaminants were present following mannose-Sepharose purification the second anion exchange purification step was abandoned. This, along with extensive dialysis into water before lyophilisation and transfer in to NMR buffer, produced high quality NMR samples. Despite extensive dialysis, residual EDTA from the mannose-Sepharose purification remains. However, EDTA can be excluded from the mannose-Sepharose purification by using a low pH elution buffer resulting in un-contaminated NMR samples (discussed further in Chapter 6).

The modifications described in this chapter allow the production of sufficient quantities of $^{13}C/^{15}N$ labelled DC-SIGNR CRD of sufficient quality for NMR analysis via triple resonance experiments, as described in further chapters.

4 Backbone Assignment

4.1 Introduction.

This chapter describes difficulties in the sequential assignment of apo (calcium free) DC-SIGNR CRD due to variable intensity of signals in the HSQC spectrum. This variable intensity is caused by intermediate dynamics on the NMR timescale (Lian & Roberts, 2011). Altering the rate of dynamics of the protein backbone by the addition of calcium and increasing the temperature overcame this problem allowing successful assignment of the holo (calcium bound) CRD. The assignment of the calcium bound protein is a useful tool for studying changes in conformation and dynamics of sugar binding (this will be discussed in chapter 7).

4.2 Variable Intensity of Apo DC-SIGNR CRD.

Despite the acquisition of the full complement of triple resonance experiments (see Materials and Methods) on the apo (calcium free) CRD only ~33% of the protein backbone was able to be assigned. This is a result of weak/absent signals in the HN(CA)CO spectrum preventing sequential

assignment. Figure 4.2-1 (A) highlights the presence of many weak signals in the HSQC spectrum of apo DC-SIGNR; this is a problem for sequential assignment as 3D experiments are inherently less sensitive than the HSQC. In particular, the HN(CA)CO is the least sensitive of the triple resonance experiments (due to rapid relaxation of the $^{13}C\alpha$ magnetization (Cavanagh, 2007)) used for sequential backbone assignment, and so any signals with weak intensity in the HSQC will likely be absent in the HN(CA)CO (Figure 4.2-2). This hinders sequential assignment substantially. The non-uniformity of the signal intensities throughout the HSQC spectrum is characteristic of a protein with varying dynamics, as the regions of the protein which display intermediate dynamics on the NMR time scale give rise to broad, low intensity signals while other regions of the protein which are in fast exchange produce stronger signal intensities. The fact that different regions of the protein backbone are in different dynamics regimes suggests that the DC-SIGNR CRD is extremely mobile in the absence of calcium. In an effort to improve the signal intensity of the HSQC spectrum, we sought to change the rate of dynamics of the protein backbone.

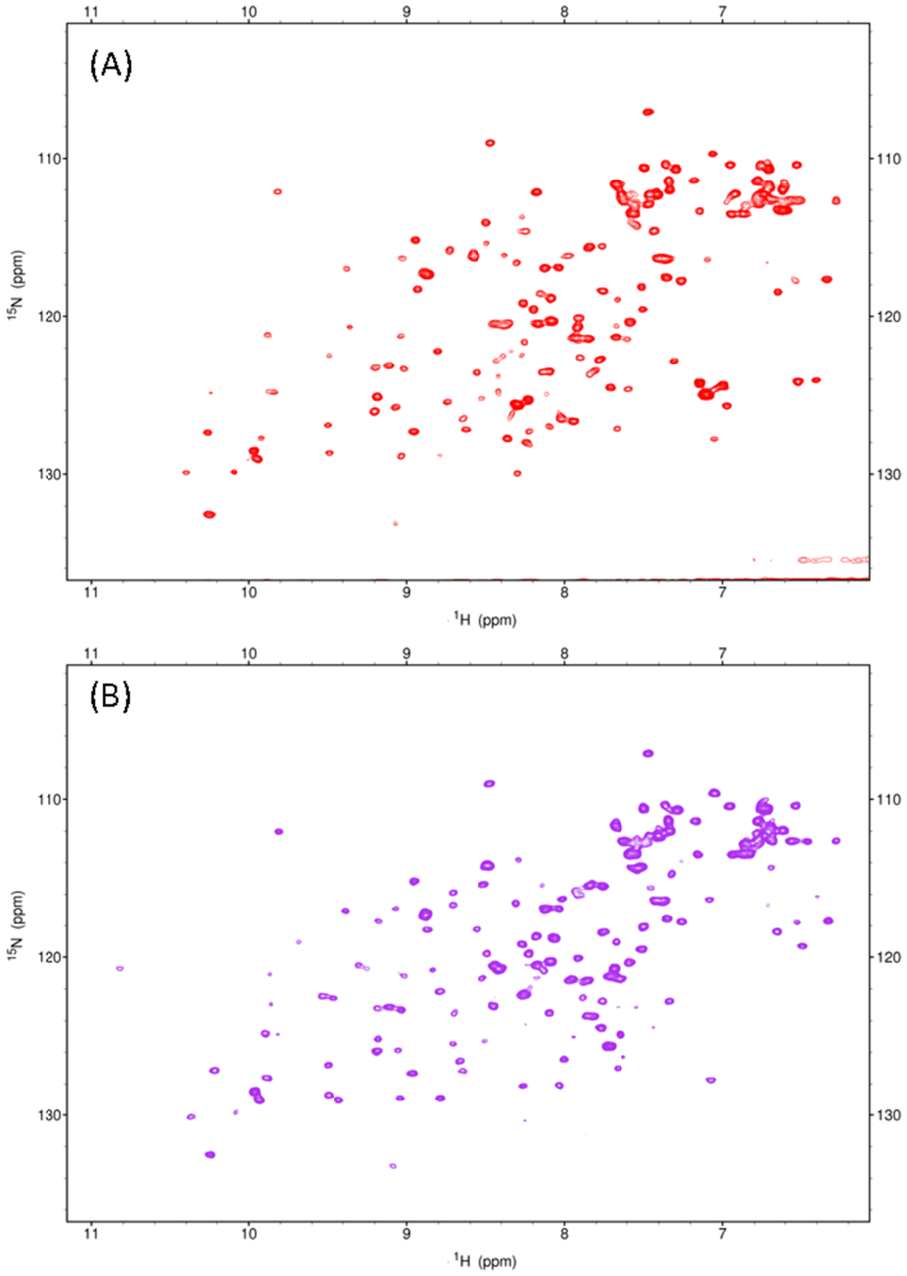

Figure 4.2-1: Comparison of ^{1}H-^{15}N HSQC spectra of (A) apo and (B) holo DC-SIGNR CRD at 25°C. The addition of 4 mM $CaCl_2$ to the DC-SIGNR CRD (B) provides a modest improvement to the variable intensity seen for the apo CRD (A). However, there is still a significant amount of variable intensity following addition of calcium indicating that the CRD remains flexible and that assignment of the protein under these conditions will not be possible.

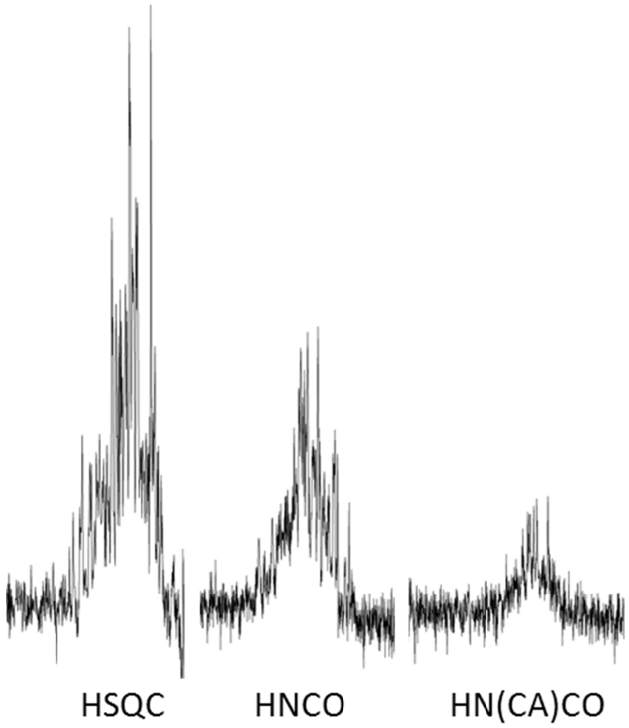

Figure 4.2-2: A comparison of the sensitivity of 2D and 3D experiments. 1D traces of 3D experiments recorded on $^{13}C/^{15}N$ DC-SIGNR CRD. The HNCO is the most sensitive of the triple resonance experiments while the HN(CA)CO is the least sensitive. These 1D traces clearly show that triple resonance experiments are far less sensitive than the HSQC experiment and that the HN(CA)CO is particularly insensitive. Therefore, it is clear that a weak intensity signal in the HSQC will not be present in the HN(CA)CO.

4.2.1 Addition of Calcium.

The binding of Ca^{2+} ions to the DC-SIGNR CRD is a requirement for glycan binding and could induce changes in the conformation and/or dynamics of the protein. Figure 4.2-1 (B) shows that the addition of 4 mM $CaCl_2$ affords a modest improvement in variable intensity throughout the HSQC spectrum. However, the addition of calcium alone is not sufficient to permit sequential assignment as evidence of variable intensity remains. This suggests that the protein maintains a high degree of flexibility following calcium binding. Interestingly, Figure 4.2-1 shows significant chemical shift changes throughout the HSQC upon addition of

calcium confirming that conformational and/or dynamics changes are taking place as a result of calcium binding. This will be discussed further in section 4.4.1.

4.2.2 Increasing Field Strength.

In an effort to improve the quality of the HSQC spectrum further, a higher field spectrometer was utilised. Figure 4.2-3 compares a HSQC spectrum of the holo CRD acquired on a Bruker 700 MHz instrument at Warwick (Figure 4.2-3 (A), purple) and a HSQC spectrum acquired on a Varian 900 MHz instrument at the university of Birmingham HWB-NMR national facility (Figure 4.2-3 (B), black). Both spectra were recorded on the same sample with the same number of scans. While, the dispersion of the spectrum acquired on the 900 MHz spectrometer is superior (as expected) the intensity is comparatively poor. This could simply be a result of variations in the sensitivity of cryoprobes produced by different manufacturers. Alternatively, increasing the field strength causes faster relaxation (as chemical shift anisotropy increases with field strength) artificially slowing the rate of protein dynamics (Prestegard, 1998) and causing more peak broadening. If this is the case, increasing the rate of dynamics could improve the variable intensity by bringing the protein backbone in to fast exchange on the NMR timescale.

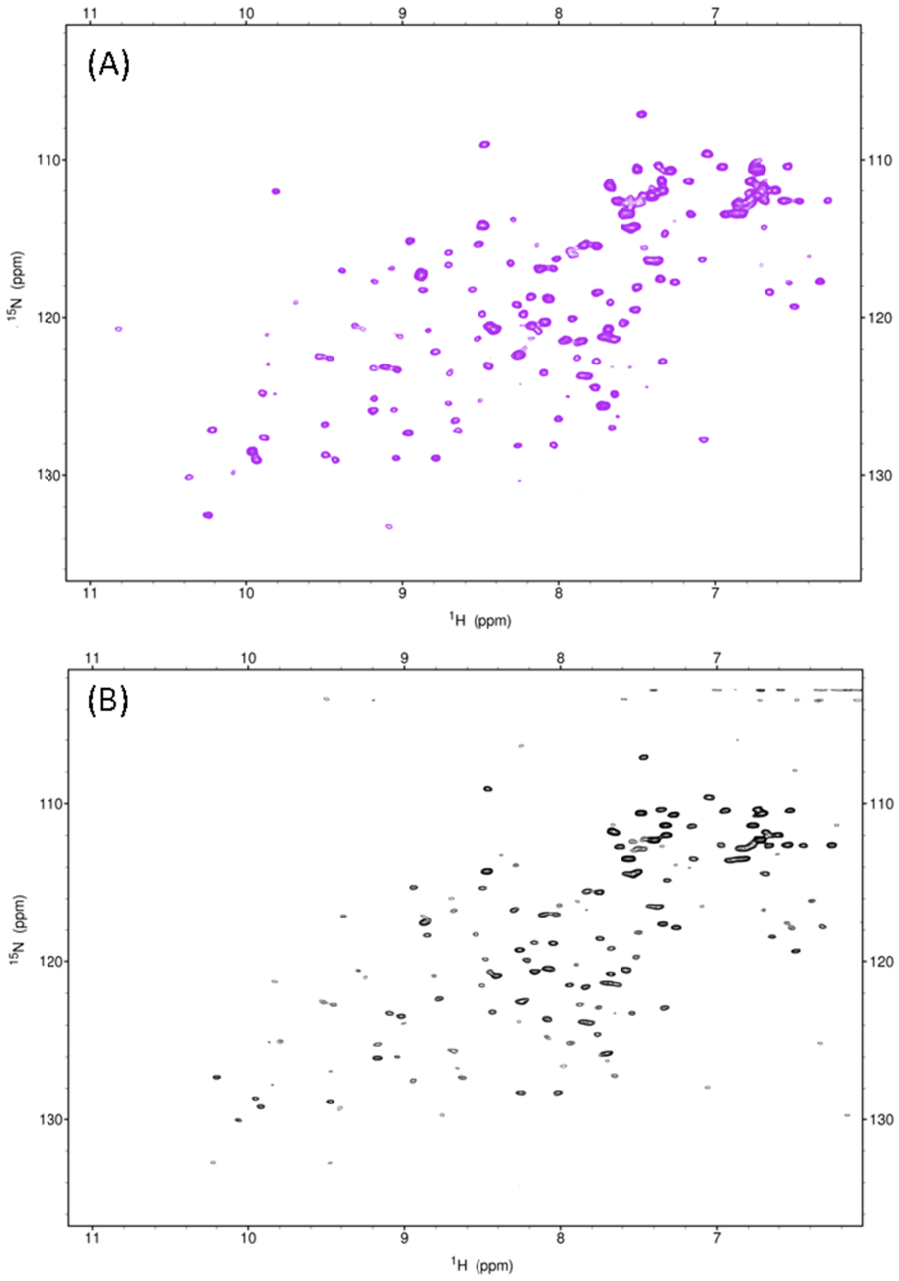

Figure 4.2-3: Comparison of ^{1}H-^{15}N HSQC spectra of holo DC-SIGNR CRD acquired at (A) 700 MHz and (B) 900 MHz at 25°C. The HSQC acquired on the Bruker 700 MHz spectrometer (purple) shows far more sensitivity than the HSQC acquired on the Varian 900 MHz spectrometer (black). Both HSQC spectra were collected on the same sample (~0.7 mM [U-^{13}C-^{15}N]-DC-SIGNR CRD, 4mM $CaCl_2$, 20mM d-Hepes, 20mM NaCl, pH 6.8) with the same number of scans.

4.2.3 Increasing Temperature.

An attempt to accelerate the rate of dynamics was made by acquiring HSQC spectra at increasing temperatures. A range of temperatures from 25°C to 37°C were tested and the most significant improvement in signal intensity was observed at 37°C. Figure 4.2-4 shows that increasing the temperature from 25°C to 37°C led to HSQC peaks which were more uniform in intensity. This, in turn, led to a significant improvement in the triple resonance experiments. Table 4.2-1 compares the number of peaks present in the most sensitive 3D experiment, the HNCO, and the least sensitive experiment, the HN(CA)CO, at 25°C and 37°C. There is a small improvement in the number of peaks present in the HNCO at 37°C (although even in this experiment some peaks are absent) while there is a much more significant improvement in the HN(CA)CO at 37°C.

Table 4.2-1: Percentage of peaks present in HNCO and HN(CA)CO spectra at 25°C and 37°C. The HNCO experiment is the most sensitive of the triple resonance experiments used for sequential assignment of the protein backbone while the HN(CA)CO is the least sensitive. At 37°C 79% of the expected signals are present in the HN(CA)CO spectrum indicating that sequential assignment will be possible under these condition.

	Visible peaks	
Experiment	25°C	37°C
HNCO	115 (86%)	121 (90%)
HN(CA)CO	65 (49%)	106 (79%)

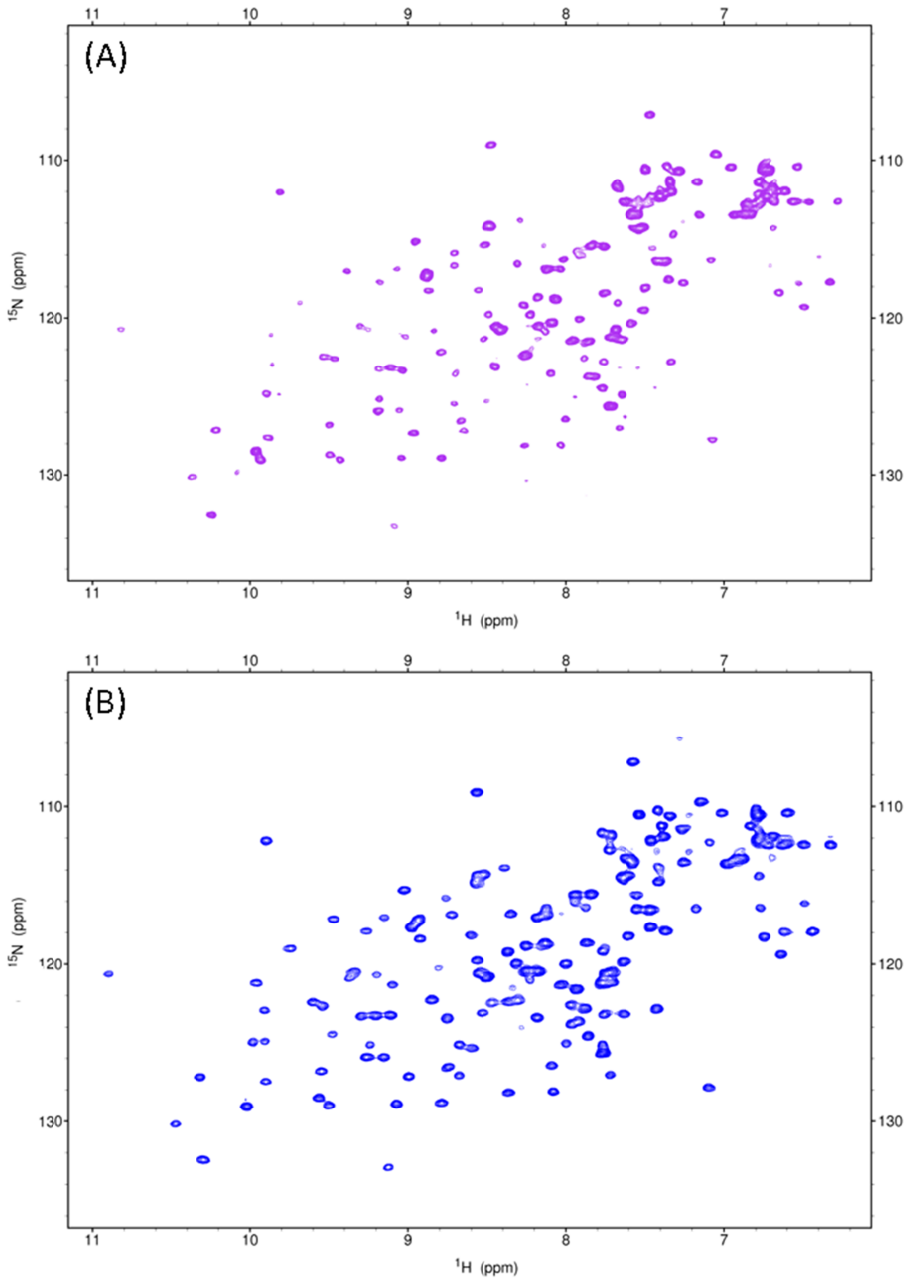

Figure 4.2-4: Comparison of ^{1}H-^{15}N HSQC spectra of holo DC-SIGNR CRD recorded at (A) 25°C and (B) 37°C at 700 MHz. A range of temperatures were investigated with the most noticeable improvement in variable intensity occurring at 37°C (B, blue). It is clear that at 37°C the intensity of the signals is more uniform.

As the greatest improvement in variable intensity was achieved with addition of calcium at 37°C on the 700 MHz spectrometer (and because calcium binding is a requirement for the formation of DC-SIGNR-oligosaccharide complexes, the study of which is a major aim of this project) these conditions were used to acquire a full suite of triple resonance experiments.

4.3 Assignment of Holo DC-SIGNR CRD.

4.3.1 An Example of the Sequential Assignment Approach.

The modifications described above produced 3D experiments of sufficient quality to allow successful assignment of the holo CRD. Figure 4.3-1 shows a chain of five sequentially assigned amino acids to highlight the quality of the 3D experiments. The HNCA/HN(CO)CA and HNCO/HN(CA)CO experiments have high resolution and are useful for making the links required to identify chains of sequential amino acids. Despite the lack of a HN(CA)CO signal in lane 1 of Figure 4.3-1 (B) the HNCA/HN(CO)CA (Figure 4.3-1 (A)) and CBCANH/CBCA(CO)NH (Figure 4.3-1 (C)) spectra confirm that this residue is part of the chain. As there is little variation in C_α and CO chemical shifts for the 20 naturally occurring amino acids, the amino acid type could not be identified by the HNCA/HN(CO)CA and HNCO/HN(CA)CO alone. As a result, the CBCANH spectrum is invaluable since the C_β chemical shift can be used to determine amino acid type (see §1.5 for more detail). We can see from the CBCANH spectrum (purple C_α and orange C_β signals) in Figure 4.3-1 (C) that strips 1 and 3 have high C_β chemical shift values which are characteristic of serine and threonine residues while strip 4 has an extremely low

C_β chemical shift value indicating that it is an alanine residue (see §1.5.2, Figure 1.5-3 for average C_β chemical shifts). As there are no serine residues followed by an alanine in the CRD sequence this chain must be $S/Tx\mathrm{TA}x$ which means there is only one possible solution: $S_{292}V_{293}T_{294}A_{295}C_{296}$. Using the triple resonance experiments in this way, the majority of the protein backbone was successfully assigned.

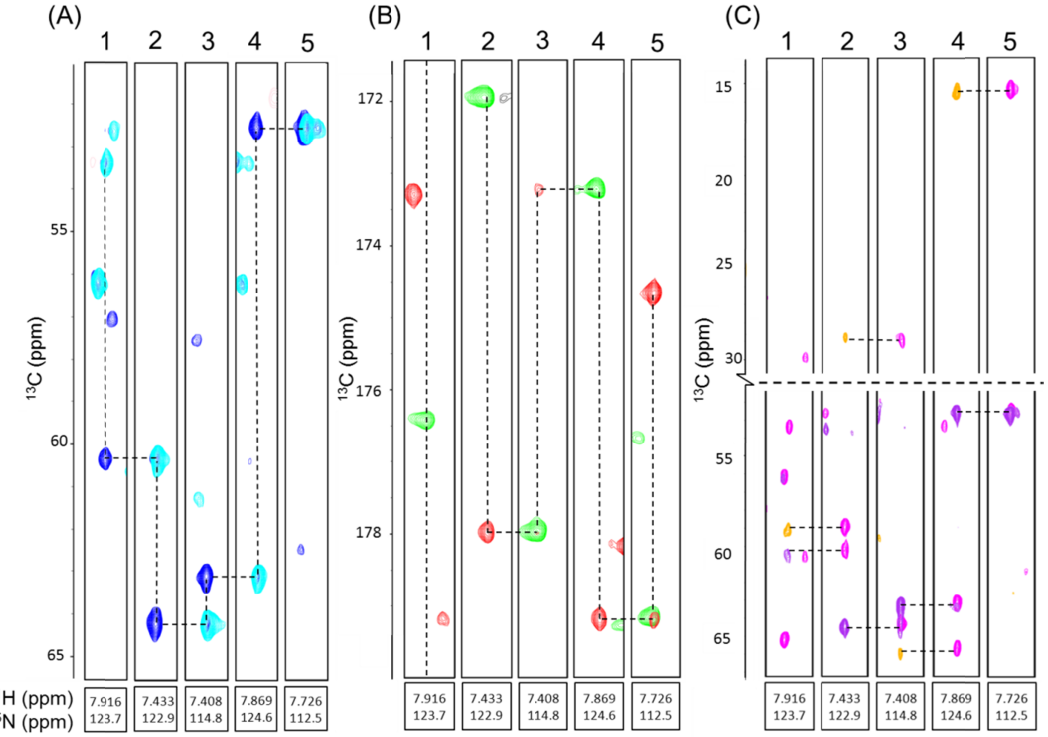

Figure 4.3-1: An example of the sequential backbone assignment of holo DC-SIGNR at 37°C. (A) HNCA (dark blue), HN(CO)CA (light blue), (B) HNCO (green), HN(CA)CO (red), (C) CBCANH (purple positive contours and orange negative contours), and CBCA(CO)NH (pink) spectra are combined to sequentially assign chains of amino acids. The sensitivity of the HNCA and HNCO experiments make them useful for sequential assignment while the C_β chemical shift from the CBCANH experiment provides valuable information on amino acid type.

4.3.2 Backbone Assignment of Holo DC-SIGNR CRD.

The assigned ^{1}H-^{15}N HSQC spectrum of the calcium-bound DC-SIGNR CRD is shown in Figure 4.3-2 and the full list of assignments is given in Appendix A,

Table A1. Of the 139 amino acids in the CRD, 5 proline residues do not appear in the HSQC spectrum due their lack of a backbone amide group and 17 residues (8 at the N terminus and 9 at the C terminus) were not present due to the flexibility of these regions (this is common in HSQCs of protein fragments as the flexible 'tail' regions sample multiple environments). Of the observable residues present in the HSQC, 96% have been unambiguously assigned (80% of the total CRD).

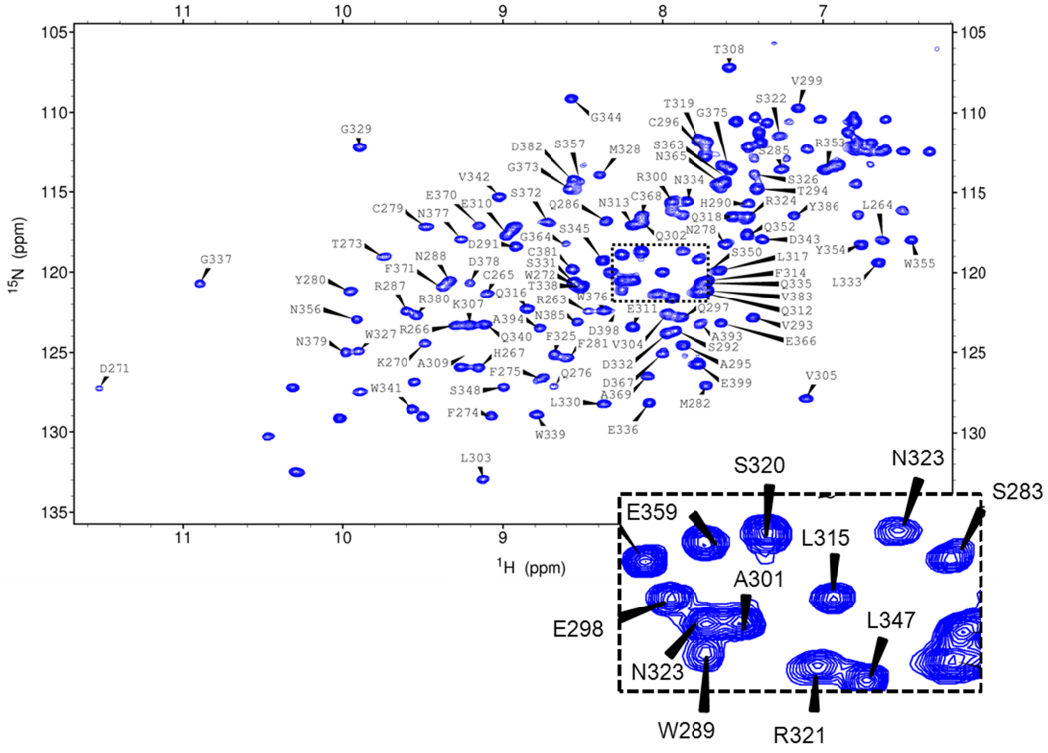

Figure 4.3-2: Assignment of holo DC-SIGNR CRD at 37°C. Addition of calcium and increasing the temperature to 37°C improved the quality of the HSQC and the triple resonance experiments sufficiently to facilitate assignment of the protein backbone. In total, 96% of the backbone residues present in the HSQC are assigned. The boxed region is enlarged to avoid overlap of labels.

The HSQC spectrum shown in Figure 4.3-2 provides a 'map' of the CRD in the calcium bound state. Since the backbone chemical shifts are sensitive to changes in chemical environment for a given nucleus, either through a direct binding event, conformational change, or change in dynamics, the assigned HSQC

spectrum is a powerful tool for probing ligand binding and allows rapid investigation of a range of ligands (this will be discussed in detail in Chapter 7).

4.3.3 Improving Assignment Using 'Sparse Unlabelling'.

Even at the increased temperature of 37°C, some residues remain un-assigned due to spectral overlap or weak intensity. In particular, residues N_{361} and N_{362} in the primary calcium binding loop (which form part of the conserved EPN motif discussed previously) cannot be assigned using the sequential assignment procedure. This indicates that the primary calcium binding site is highly dynamic even in the presence of calcium (the dynamics of the calcium binding loops is discussed further in §7.5). As these residues are of particular interest due to their location in the calcium binding loop, a sparse unlabelling approach was used to facilitate their assignment (as described in Materials and Methods §2.6.5).

The unlabelling approach aims to identify all signals in the HSQC spectrum which arise from a specific amino acid type by eliminating their signals. This is achieved by the addition of the unlabelled amino acid of interest to ^{15}N labelled M9 media prior to induction. If the amino acid is incorporated in to the protein successfully, all signals arising from that amino acid will be absent from the HSQC spectrum. Comparison with a HSQC spectrum of uniformly labelled protein allows fast identification of the signals corresponding to the amino acid type of interest. However, metabolic scrambling can result in unlabelled atoms being propagated throughout the protein (Hiroaki et al, 2011) making the technique unreliable. To address this, an amino acid with no ambiguous assignments (threonine) was used. Figure 4.3-3 confirms that the amino acid of interest was successfully

'unlabelled' since the only signals absent from the spectrum (Figure 4.3-3, orange) correspond to the, already assigned, threonine residues. This verifies that there is no metabolic scrambling in this system and that the unlabelling strategy can be utilised to identify un-assigned amino acids. In addition, Figure 4.3-3 confirms that the assignments of all threonine residues made using the sequential assignment technique are correct. However, it is worth noting that the absence of residue T_{388} is not clear in Figure 4.3-3 indicating that this technique is not useful in cases of significant signal overlap.

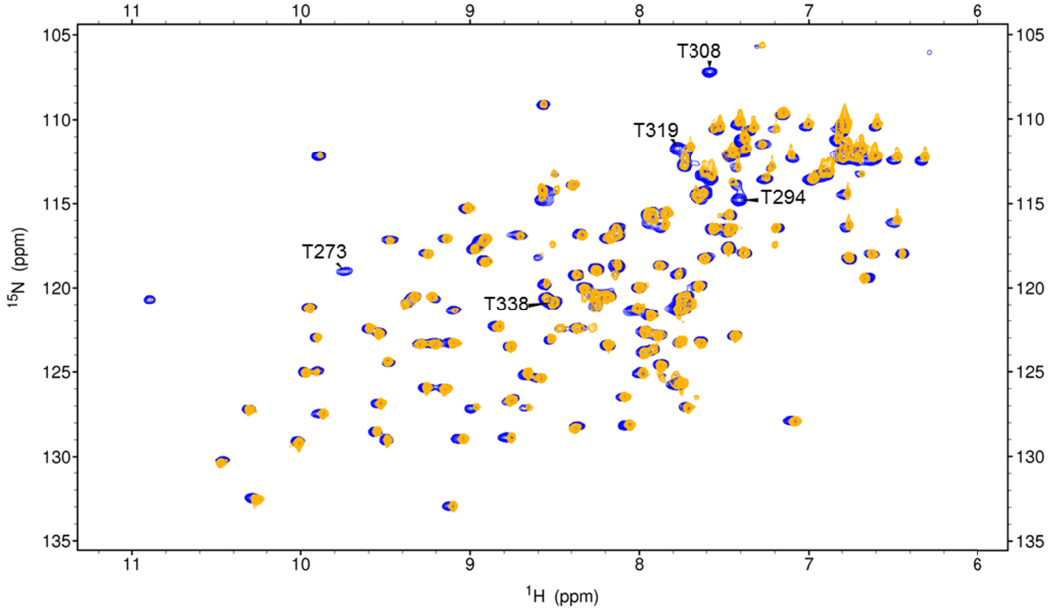

Figure 4.3-3: Overlay of [U-^{15}N]-CRD and T-unlabelled CRD. As expected the spectra of unlabelled threonine residues (orange) contains less peaks than the uniformly labelled spectra (blue). All residues that are missing in the unlabelled spectrum correspond to the Thr residues confirming that this is a viable technique for identification of difficult to assign amino acids. However, residue T_{388} is not clearly absent showing that this technique is not useful when there is significant overlap.

The T-unlabelled sample in Figure 4.3-3 confirmed the viability of the unlabelling technique, therefore an N-unlabelled sample was prepared in an effort to identify the missing assignments in the primary Ca^{2+} binding loop. Figure

4.3-4 shows an overlay of the HSQC spectra recorded on the unlabelled sample (green) and the uniformly labelled sample (blue). All of the assigned Asn residues were absent from the unlabelled spectrum (confirming that these assignments are correct), and several additional peaks were also missing. The signals circled in Figure 4.3-4 are due to the asparagine side-chain amides, while the two missing signals at ~117 ppm (inset panel in Figure 4.3-4) are due to the backbone amides of the missing assignments N_{361} and N_{362}. It is worth noting that residues K_{270} and G_{337} are extremely weak in the unlabelled HSQC (Figure 4.3-4 arrows). However, these residues often give low intensity signals at low protein concentrations and at higher field and so this is likely due to the lower concentration of the asparagine unlabelled protein and not metabolic scrambling. Although the unlabelled strategy allows us to identify where the signals corresponding to residues N_{361} and N_{362} appear in the HSQC spectrum, we cannot distinguish between the residues because they are absent in the 3D spectra. This is almost certainly due to extremely fast ^{15}N relaxation (see §7.5 for dynamics data). Nevertheless, this ambiguous assignment allows us to record chemical shift perturbation and dynamics data of this region of the primary calcium binding loop.

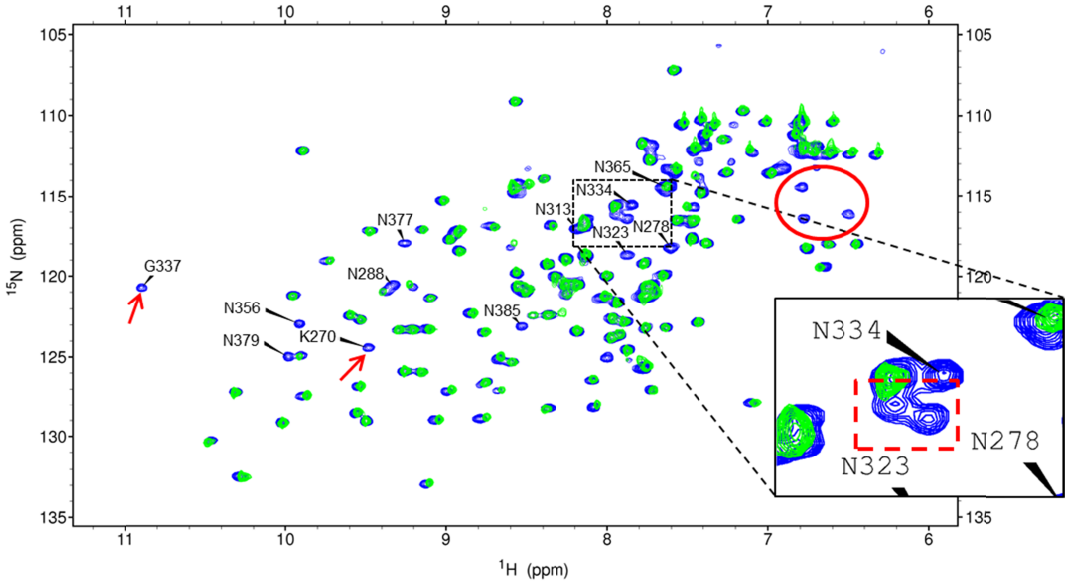

Figure 4.3-4: Overlay of [U-^{15}N]-CRD and N-unlabelled CRD. All assigned asparagine residues are absent in the spectra of N unlabelled DC-SIGNR (green) along with some additional residues. The missing residues at low ^{1}H ppm (circled) are due to asparagine side-chains while the missing signals at ~117 ppm circled in the inset panel are due to the missing assignments (residues 103 and 104).

4.4 Assignment of Apo DC-SIGNR CRD.

Although the complete assignment of apo DC-SIGNR using triple resonance experiments was not possible due to the variable intensity of the HSQC spectrum, the assignment of the holo CRD can be used to infer information on calcium binding. Calcium binding to the DC-SIGNR CRD has not been studied in detail and so any information that can be gained from a comparison of the holo and apo HSQC spectra is of interest.

Despite the large differences between the spectra, the assignment of the holo DC-SIGNR spectrum can be superimposed on to the apo spectrum allowing a tentative assignment of the un-perturbed residues. Peaks that were not perturbed by addition of calcium and isolated peaks (no other chemical shifts nearby) where there was no ambiguity as to the corresponding peak in the holo

HSQC spectrum were assigned in the apo HSQC. This tentative assignment was combined with information from the triple resonance experiments (chains of sequential amino acids and partial assignments) to improve the assignment of the apo CRD (Figure 4.4-1). Using the holo assignment as a guide, ~63% of the residues in the apo DC-SIGNR CRD were assigned (an improvement from the ~33% assigned using triple resonance experiments alone). This tentative assignment facilitated a preliminary investigation of calcium binding.

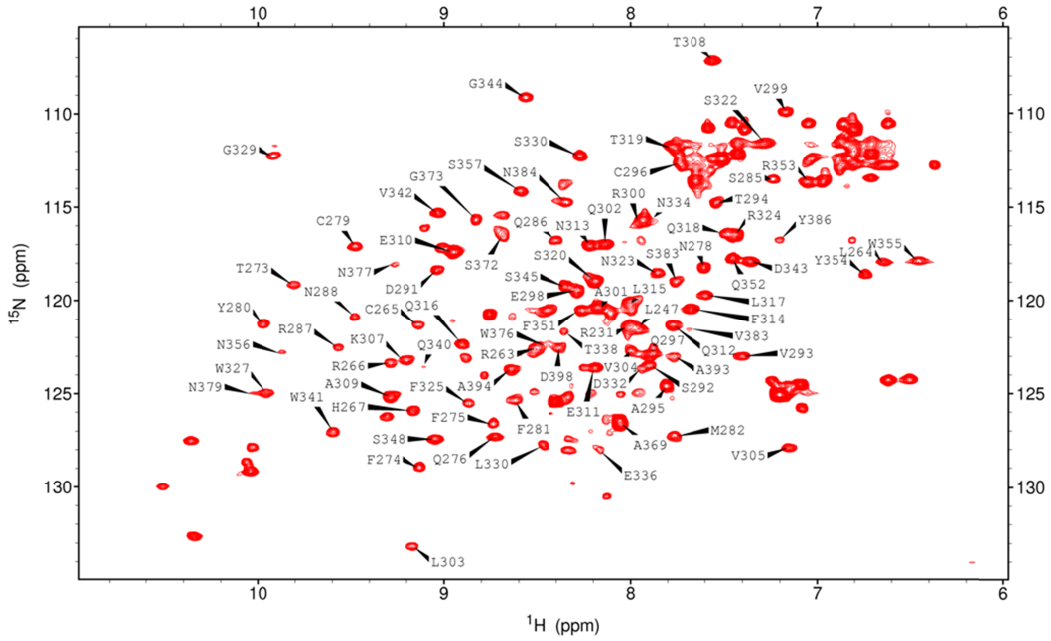

Figure 4.4-1: Partial assignment of apo DC-SIGNR CRD. Superimposing the chemical shift assignments from the holo HSQC on to the apo combined with sequential assignment using triple resonance experiments resulted in assignment of ~63% of the apo protein backbone.

4.4.1 Calcium Binding Induces Significant Conformational and Dynamics Changes Throughout the CRD.

Figure 4.4-2 highlights that significant chemical shift perturbations occur throughout the HSQC upon calcium binding. Almost every peak in the spectrum is perturbed indicating that significant changes are occurring throughout the CRD. Several amino acids distal to the calcium binding loops are significantly perturbed

(for example F_{274} is located at the N-terminal and T_{294} in α-helix 1, both regions which are not implicated in calcium binding) further highlighting that calcium addition causes global changes to the CRD. These conformational and/or dynamics changes could explain the increased stability of the holo CRD and could result in formation of the ligand binding site. Indeed, residues F_{325} and S_{320} which form part of the binding site in the crystal structure (the 'shelf' formed by α-helix 2) are perturbed substantially.

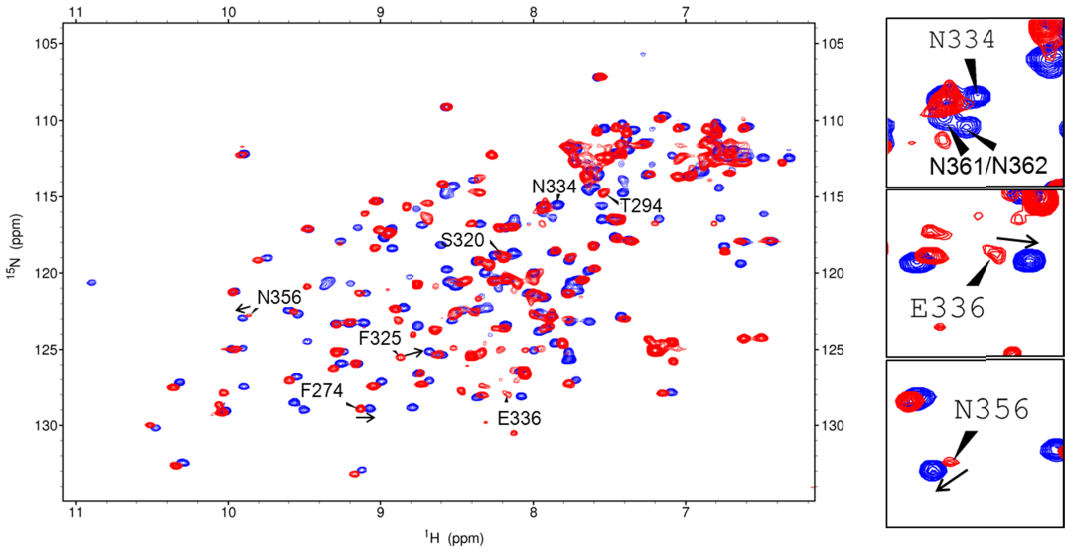

Figure 4.4-2: Overlay of apo (red) and holo (blue) HSQC spectra at 37°C. Although recording the HSQC spectra of apo DC-SIGNR (red) at 37°C improves the variable intensity the quality is not as good as the holo HSQC (blue). By overlaying the two spectra we can see there are a great deal of chemical shift perturbations as a result of calcium binding. This indicates that a significant conformational and/or dynamics change occurs throughout the CRD. A few interesting residues, that undergo significant changes as a result of calcium binding, are highlighted.

In the HSQC spectrum of apo DC-SIGNR residues that correspond to the primary calcium binding site cannot be assigned either due to absence of the signal or substantial chemical shift perturbations from values observed in the holo-form. Residues N_{361} and N_{362} in the primary calcium loop are absent in the apo HSQC spectrum. This could be a result of *cis-trans* isomerisation of residue

P_{360} (in the conserved EPN motif of calcium binding loop 1) in the absence of calcium (this has been shown for other C-type lectins) (Ho et al, 2010; Ng et al, 1998; Nielbo et al, 2004; Pavlicek et al, 2003; Poget et al, 2001). In addition, residue N_{334} in the secondary calcium loop is not visible in the apo HSQC, and residues N_{356}, at the beginning of the primary loop, and E_{336}, which is implicated in direct Ca^{2+} binding in the secondary loop, are significantly perturbed and have extremely weak intensity in the apo spectrum. This indicates that the dynamics of the calcium binding loops is altered substantially in the presence of calcium.

4.5 Conclusion.

Analysis of HSQC spectra of both the apo and holo DC-SIGNR CRD highlights the dynamic nature of this C-type lectin. The variable intensity at room temperature indicates that the CRD is mobile while the significant chemical shift perturbations and line width changes in the holo HSQC are characteristic of binding induced conformational and dynamics changes. The perturbations occur throughout the HSQC suggesting a global change in the CRD on calcium binding which could result in formation of the ligand binding site. In addition, residues which reside in the calcium binding loops substantially increase in intensity in the holo HSQC which is indicative of a dynamics change in these regions. The mobility of the protein may play a functional role by increasing affinity for glycans which are also inherently flexible and structural/dynamics changes as a result of calcium may explain the lack of glycan binding in the absence of calcium. Although addition of calcium improves the intensity of several HSQC signals there is still a

degree of variable intensity at 25°C suggesting that the protein maintains a level of mobility in the calcium bound state. However, this was improved by increasing the temperature to 37°C allowing a nearly complete assignment of the holo CRD backbone residues. This is the first reported solution state NMR assignment of the DC-SIGNR CRD. The assigned HSQC is a powerful tool for the study of ligand interactions which will be described in Chapter 7.

5 Side-chain Assignment and Work Towards a Solution Structure.

5.1 Introduction.

Solution structure calculations require measurement of nuclear Overhauser effects (NOEs). This transfer occurs through-space and is a powerful tool for determining distance restraints which allow the full 3D tertiary structure of proteins to be calculated. In order to determine which nuclei are experiencing NOE transfer, a full side-chain assignment (assignment of all carbons and protons in the DC-SIGNR CRD) is required. This assignment is used to identify the NOE peaks and then calculate distance constraints (based on the intensity of the NOE signals) and perform a structure calculation (these steps are usually carried out iteratively using a programme such as CYANA). In addition to the side-chain assignment and NOE peak intensities, the structure calculation also requires

information on phi and psi torsion angles. The torsion angles are calculated from prediction of the protein secondary structure using the chemical shift index method.

This chapter reports the results of predicting the secondary structure of the DC-SIGNR CRD using the backbone assignments described in §4.3.2 along with the H_α assignments obtained from experiments described below. In addition, the experiments required for the full side-chain assignment of the CRD are described and the full table of chemical shifts is reported in Appendix A. The full side-chain assignment of the CRD paves the way for solution structure elucidation via $^{15}N/^{13}C$-NOESY-HSQC experiments. Appropriate mixing times for these experiments have been determined and full data sets acquired. Unfortunately, structure calculations using the full assignment, torsion angle constraints and NOE intensities from the ^{15}N-NOESY-HSQC spectrum did not converge. As a result, ^{13}C-NOESY-HSQC spectra have been acquired because this experiment contains more information on side-chain distance restraints. Work to solve the structure with addition of the ^{13}C-NOESY-HSQC restraints is on-going.

5.2 Secondary Structure Prediction is in Good Agreement with the Crystal Structure.

The assignment of the backbone residues (C_α, C_β, CO) described in §4.3.2 (along with the H_α assignments obtained from the side-chain experiments described below in §5.3) allows a prediction of the secondary structure elements present in the CRD to be made using the chemical shift index method (Wishart &

Sykes, 1994; Wishart et al, 1992). This method is based on the observation that the chemical shifts of residues in defined secondary structure elements are shifted from the average (random coil) values. For example, H_α, C_α and CO are shifted to higher chemical shift values (deshielded) when in α-helices while the C_β is shifted to lower chemical shift values (shielded). In this case the amino acid is assigned the value -1. The opposite is true for β-sheet regions (amino acids with these chemical shifts are assigned the value +1). If four or more consecutive amino acids are assigned the value -1, then that region is classified as an α-helix, while three or more consecutive amino acids with the value +1 are required to define a β-sheet region. Using this method, the secondary structure of the protein can be estimated. In addition, the programme TALOS+ can be used to predict the dihedral angles which in turn are used to predict secondary structure elements. The phi and psi torsion angles predicted by TALOS+ can be used as input in solution structure calculation programmes such as CYANA (Guntert, 2004).

Figure 5.2-1 shows the secondary structure prediction per residue for the DC-SIGNR CRD using the programme TALOS+ (Shen et al, 2009) compared to the secondary structure in the 1K9J crystal structure. There is good agreement with the crystal structure for the N-terminal section of the CRD. However, TALOS+ makes no predictions in the C-terminal end of the CRD likely due to missing assignments in this region.

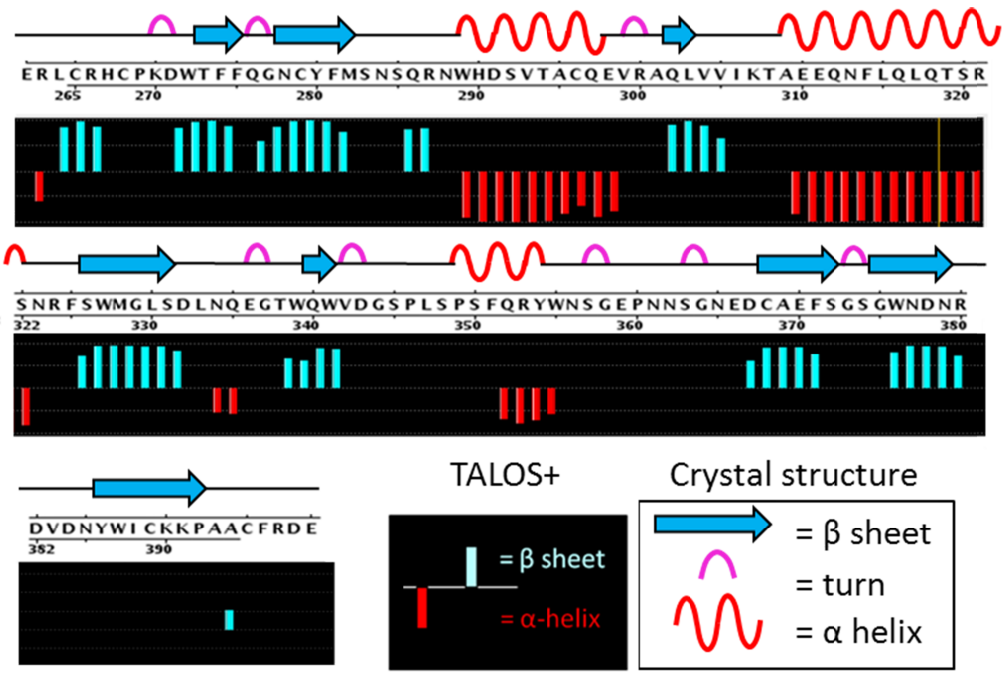

Figure 5.2-1: Secondary structure prediction. The C_α, C_β, CO and H_α chemical shifts were used to predict the dihedral angles and, in turn, the secondary structure elements of the CRD in solution using the programme TALOS+. There is good agreement with the secondary structure in the crystal for the N-terminal portion of the CRD. However, TALOS+ cannot make predictions for the C-terminal region likely due to the presence of un-assigned residues.

5.3 Side-Chain Assignment.

In order to complete the assignment of the DC-SIGNR CRD, experiments which correlate all side-chain protons and carbons are required. The TOCSY-HSQC experiment (Marion et al, 1989b) is the logical choice for this since the ^{1}H-^{15}N-HSQC spectrum is extended into a third TOCSY dimension (containing through-bond proton couplings) resulting in the correlation of each backbone amide with every proton in the amino acid. However, this experiment is, in general, not used for proteins with molecular weights >15 KDa (Bertini et al, 2012). Indeed, the HSQC-TOCSY spectrum of the DC-SIGNR CRD produced very few side-chain correlations due to ^{15}N relaxation causing loss of signal during the TOCSY mixing time. As an alternative, the H(CCO)NH (Montelione et al, 1992) experiment

(illustrated in Figure 5.3-1) was used. This experiment provides the same information as the HSQC-TOCSY, but the backbone amide is correlated with the side-chain of the previous residue and the magnetization is transferred via the side-chain carbons (Figure 5.3-1). Unlike the TOCSY experiment which uses isotropic mixing to transfer magnetization between spins (during which the magnetization is lost due to relaxation), the H(CCO)NH does not use this approach and thus can overcome the problems experienced with the HSQC-TOCSY and provide more side-chain information. Figure 5.3-1 shows that while there is signal present in the H(CCO)NH, there are a significant number of missing side-chain correlations. However, the H_α correlation was present for almost all amino acids allowing these to be quickly assigned, when using the assigned ^{1}H-^{15}N-HSQC as a guide. The H_α assignment, along with the previously assigned C_α and C_β residues, provide enough information for the use of experiments which lack the backbone amide correlation and hence are not based on the ^{1}H-^{15}N HSQC spectrum.

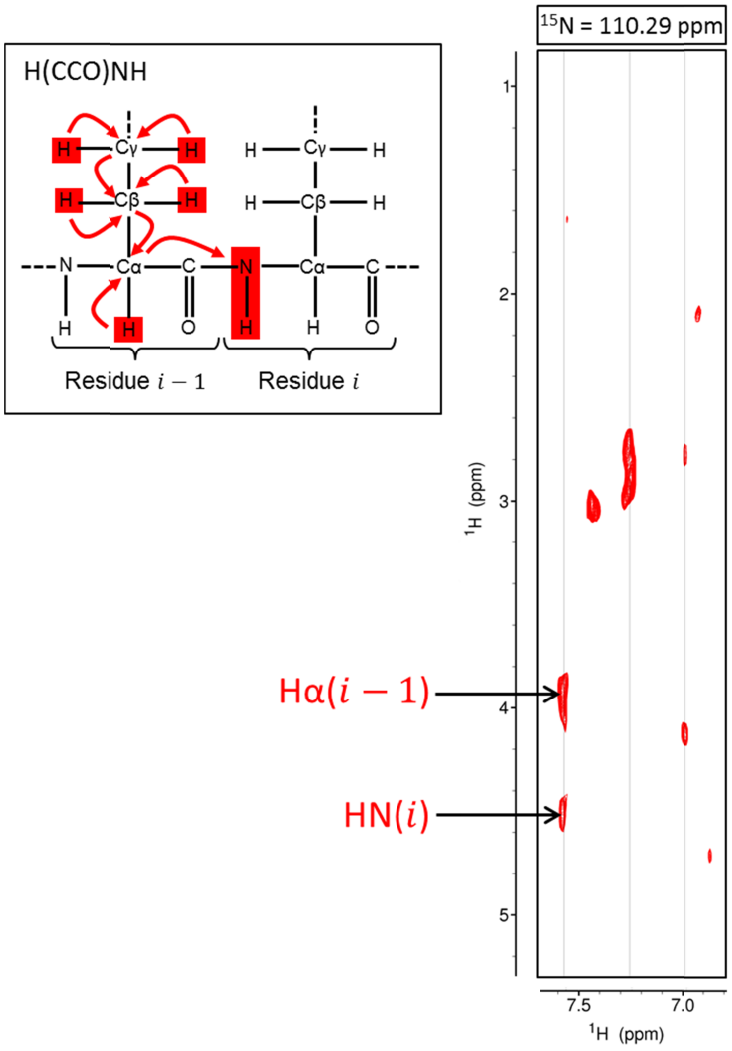

Figure 5.3-1: Example of H(CCO)NH spectrum. During the H(CCO)NH spectrum magnetization is transferred from all protons of residue $(i-1)$ to their corresponding carbon atom. The magnetization is then correlated to the backbone amide of residue (i). The result is a 3D data set with two ^{1}H dimensions and a single ^{15}N dimension. Using the assigned ^{1}H-^{15}N-HSQC this experiment should allow all side-chain protons to be identified.

Figure 5.3-2 illustrates the magnetization transfer pathways of the CCH-TOCSY (Fesik et al, 1990) and HCCH-TOCSY experiments (Bax et al, 1990). During the CCH-TOCSY, magnetization is transferred from a proton to the directly attached carbon and then to all other carbons in the side-chain. The result is a 3D data set with two carbon dimensions and a single hydrogen dimension. A single ^{13}C-^{13}C plane at proton chemical shift (x) contains a diagonal peak which

represents the directly attached carbon (the carbon bound to a proton with chemical shift x ppm). All other carbon atoms in the side-chain appear horizontally to this diagonal peak (marked with a diagonal line in Figure 5.3-2). In contrast, the HCCH-TOCSY spectrum contains two ^{1}H dimensions and a single ^{13}C dimension. A ^{1}H-^{13}C plane shows all protons in an amino acid aligned horizontally at the ^{13}C chemical shift of the directly attached carbon. For example, if the 3rd dimension (direct-^{1}H) is set to the chemical shift of the H_α nucleus in residue 20, signals for all protons in the side-chain of residue 20 will be aligned horizontally at the chemical shift of the C_α nucleus. If the 3rd dimension is then moved to the chemical shift of the H_β nucleus the same pattern of side-chain protons will now be aligned horizontally at the carbon chemical shift of the C_β nucleus. In this way the chemical shifts of all side-chain protons for each amino acid can be identified using the HCCH-TOCSY. By moving to these proton chemical shifts in the CCH-TOCSY, all side-chain carbons can be identified and the carbon directly attached to the proton is known because it falls on the diagonal.

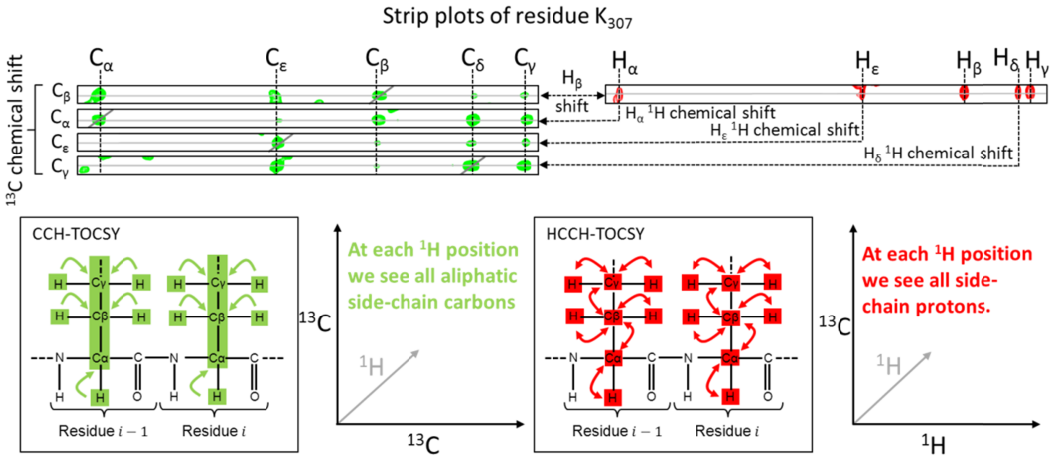

Figure 5.3-2: HCCH-TOCSY (red) and CCH-TOCSY (green) experiments**.** The CCH-TOCSY is used to assign all aliphatic side-chain carbons while the HCCH-TOCSY is used to assign all side-chain protons. If a side-chain proton is found in the HCCH-TOCSY spectrum moving to that ^{1}H chemical shift in the CCH-TOCSY will show all the side-chain carbons in that amino acid, with the carbon on the diagonal representing the carbon directly bound to the side-chain proton.

The HCCH-TOCSY and CCH-TOCSY experiments do not involve magnetization transfer via the backbone amide, so the HSQC spectrum cannot be used to guide the assignment procedure. As a result, a degree of side-chain assignment is required in order to navigate the spectrum. In this case, the C_α and C_β assignment (from the triple resonance experiments described previously in §4.3.1) along with the H_α assignment from the H(CCO)NH experiment were sufficient. The distinctive C_β chemical shift allowed amino acids to be identified in the CCH-TOCSY experiment while the C_α and H_α chemical shifts were used to verify that the correct amino acid was found. The distinctive pattern of side-chain chemical shifts of each amino acid (see Figure 5.3-3 for all carbon chemical shifts) provides further confirmation that the assignment is correct.

Figure 5.3-2 illustrates the side-chain assignment procedure utilised for the CCH-TOCSY and HCCH-TOCSY. First, an amino acid was identified by searching

for the C_β chemical shift. Residue K_{307} is used as an example in Figure 5.3-2. The chemical shifts of the already assigned C_α and H_α nuclei along with the pattern of side-chain carbon chemical shifts (see Figure 5.3-3) confirm that this is in fact residue $K_{307,}$ so the carbon peaks in the CCH-TOCSY can be assigned. As the C_β signal is on the diagonal, the corresponding proton chemical shift (in the 3rd dimension of the HCCH-TOCSY) must represent the H_β atom. Moving to this ^{1}H chemical shift in the HCCH-TOCSY allows the identification of the side-chain protons of residue K_{307} (these are aligned horizontally along the chemical shift of the C_β nucleus in the ^{13}C dimension). The side-chain protons are assigned by identifying the directly attached carbon in the CCH-TOCSY spectrum. For example, at the H_α chemical shift position in the CCH-TOCSY the same pattern of K_{307} peaks are found with the C_α peak on the diagonal. This confirms the H_α assignment. Using the CCH-TOCSY and HCCH-TOCSY, ~75% of the expected side-chain peaks have been assigned (see Appendix A for a full chemical shift assignment table).

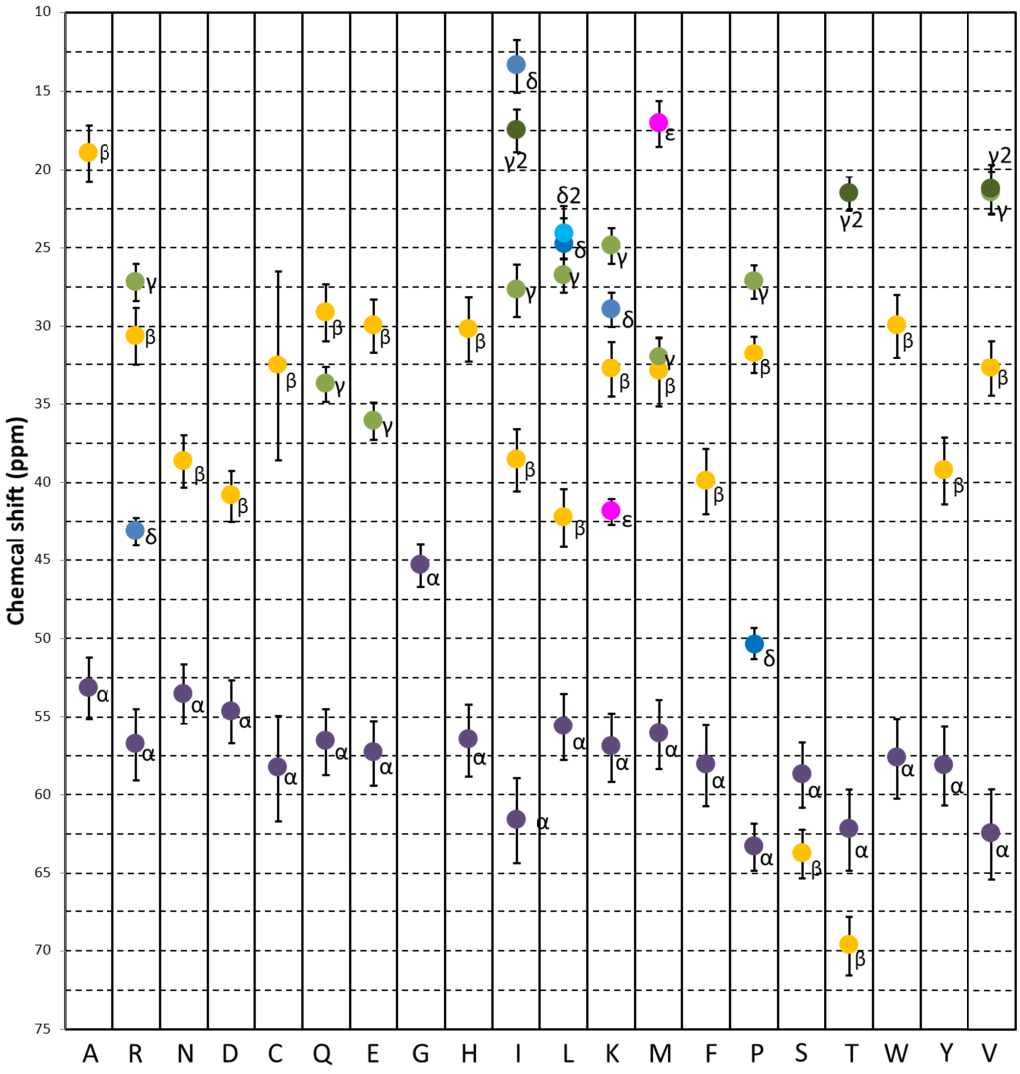

Figure 5.3-3: All average carbon chemical shifts. Average carbon chemical shifts for each of the 20 commonly occurring amino acids. Most amino acids can be identified by their distinctive side-chain carbon chemical shift patterns.

5.4 Distance Constraints.

A complete residue assignment is a prerequisite for nuclear Overhauser effect (NOE) based structural elucidation. The NOE refers to the transfer of polarization between spins as a result of dipole-dipole interactions (relaxation due to the magnetic fields produced by neighbouring nuclei) known as cross-relaxation. As a result the NOE is dependent on the distance between nuclei as well as their mobility (Williamson, 2009). The intensity of the observed NOE is

proportional to $\frac{1}{r^6}$ where r is the distance between nuclei (Williamson, 2009). This allows intermolecular distances to be calculated, thus providing information on the molecular geometry which is indispensable to the calculation of 3D protein structures. Figure 5.4-1 (A) illustrates that by increasing the amount of time that the NOE is allowed to evolve (the mixing time) then larger distances can be reported.

The $\frac{1}{r^6}$ dependence assumes that the protein is rigid and tumbles isotropically in solution. Internal motions decrease the rate of cross-relaxation, in turn decreasing the intensity of the NOE peak. However, this does not pose a significant problem as NOE distances are usually pooled into categories designated as weak, intermediate and strong as opposed to assigning an exact distance. This means that if a NOE peak has a low intensity due to protein mobility (the nuclei are close together but due to the mobility of the protein the observed NOE is weaker than expected) then it is assigned as a 'weak' restraint rather than a 'strong' restraint (very close in space). Assigning a NOE peak as a 'weaker' restraint than it should be, does not severely hinder the structure calculation procedure. However, the mobility of the protein can result in disappearance of peaks and fewer distance restraints, which may be more problematic. The largest problem in analysing NOE distance restraints arises from spin-diffusion – the transfer of NOEs along molecular bonds. As an NOE can be more efficiently transferred along bonds than through space, spin-diffusion can result in an intense NOE signal between nuclei that are distant in space but linked through one or more bonds. The result is that peaks that are distant in space are

assigned a short distance constraint ('strong'), with the knock-on effect that NOE signals representing nuclei that are close in space appear weak in comparison and so are assigned a longer ('weak') distance constraint. This can be overcome by simply relaxing the distance constraints or by measuring NOEs at short mixing times where the effect of spin-diffusion is minimized (Cavanagh, 2007).

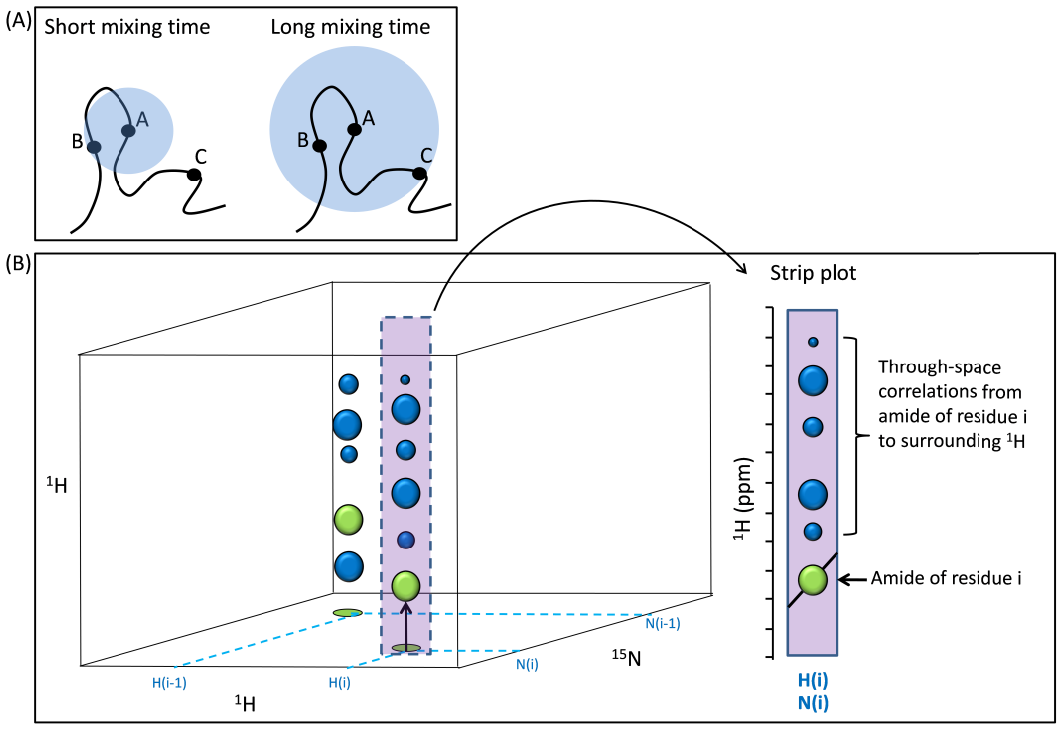

Figure 5.4-1: (A) Illustration of NOE transfer with increasing mixing time and (B) ^{15}N-NOESY-HSQC experiment. (A) As the mixing time is increased, the NOESY magnetization is transferred further through-space allowing longer distances to be observed. At a short mixing time NOEs are observed between amino acid A and B. If the mixing time is increased then larger distances are reported and a NOE is observed between amino acid A and C. (B) The ^{15}N-NOESY-HSQC extends the HSQC spectrum into a 3rd ^{1}H dimension. During this experiment magnetization is transferred through space and so a diagonal peak for the amide proton (green) is observed and protons near the amide in space are correlated vertically with the intensity of the signal decreasing with increasing distance from the amide.

5.4.1 ^{15}N NOE Distance Constraints.

The ^{15}N-NOESY-HSQC is essential for determination of distance restraints. This experiment combines a ^{1}H-^{15}N-HSQC with a NOESY (nuclear Overhauser

effect spectroscopy) sequence resulting in a 3D experiment containing two ^{1}H dimensions and a single ^{15}N dimension (Marion et al, 1989a). The data are comprised of a series of ^{1}H-^{1}H planes, each at a specific ^{15}N chemical shift, with the HN signal along the diagonal and NOESY cross peaks to nearby protons in a vertical strip (Figure 5.4-1 (B)). This allows distance constraints between the backbone amide and all nearby protons to be determined: using the side-chain assignment the NOE cross peaks are assigned distance constraints based on their relative intensities.

Multiple ^{15}N-NOESY-HSQC experiments were acquired with varying mixing times in order to determine the optimal mixing time for distance constraint determination (Figure 5.4-2). At short mixing times, the NOE can only report over a short distance and hence a small number of peaks are observed. As the mixing time increases, the NOE reports over longer distances, giving rise to more peaks. However, as the mixing time increases so do the contributions from spin-diffusion, severely hindering the determination of accurate distance constraints. Figure 5.4-2 shows that at the shortest mixing times tested (50-120 ms), there are a small number of NOE peaks. This is unexpected as large molecules experience fast NOE build-up (because long rotational correlation times promote cross-relaxation) and suggests that internal mobility is causing NOE signals to be lost. This further highlights the flexibility of the CRD and supports the variable intensity seen in the HSQC spectra at room temperature and the lack of signal in the HSQC-TOCSY experiments due to ^{15}N relaxation. The greatest number of peaks occurs with 300 ms mixing time. However, at this prolonged mixing time there is likely a significant contribution from spin-diffusion and so this mixing time

should be avoided. As a result, a mixing time of 200 ms was deemed most useful for determination of distance restraints for structure calculation (although spin-diffusion cannot be ruled out and so restraints may need to be relaxed during initial structure calculations).

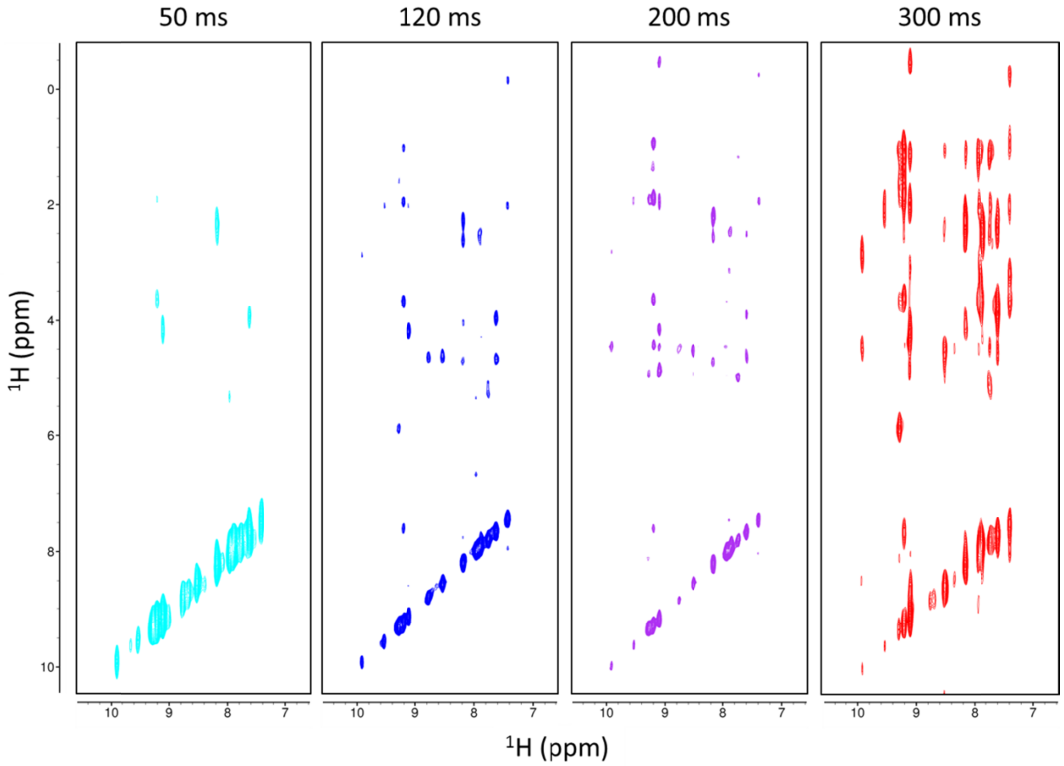

Figure 5.4-2: ^{15}N-NOESY-HSQC spectra at different mixing times. Increasing the mixing time increases the amount of cross-relaxation that occurs and hence as the mixing time increases nuclei that are further away can be seen. However, as the mixing time is increased the effects from spin-diffusion are increased and so accurate distance restraints are difficult to obtain.

5.4.2 Structure Calculation Using ^{15}N NOE Constraints Does Not Converge.

A structure calculation was attempted using the programme CYANA and distance restraints obtained from the ^{15}N-NOESY-HSQC, torsion angle restraints obtained from the secondary structure prediction using TALOS+ and the full (backbone and side-chain) assignments. The automated, combined NOESY

assignment and structure calculation method was used (Figure 5.4-3). There is the potential for a significant number of restraints from this data alone as a total of 665 NOE peaks were present in the ^{15}N-NOESY-HSQC. However, the structure did not converge because the guidelines for successful runs were not met: the number of unassigned NOEs and the number of discarded long-range NOEs were both greater than 20%, and the RMSD value of cycle 1 was greater than 3 Å (Guntert, 2004). The failure of the combined NOE assignment and structure calculation protocol to produce a satisfactory structure is possibly due to the lack of ^{13}C NOE constraints.

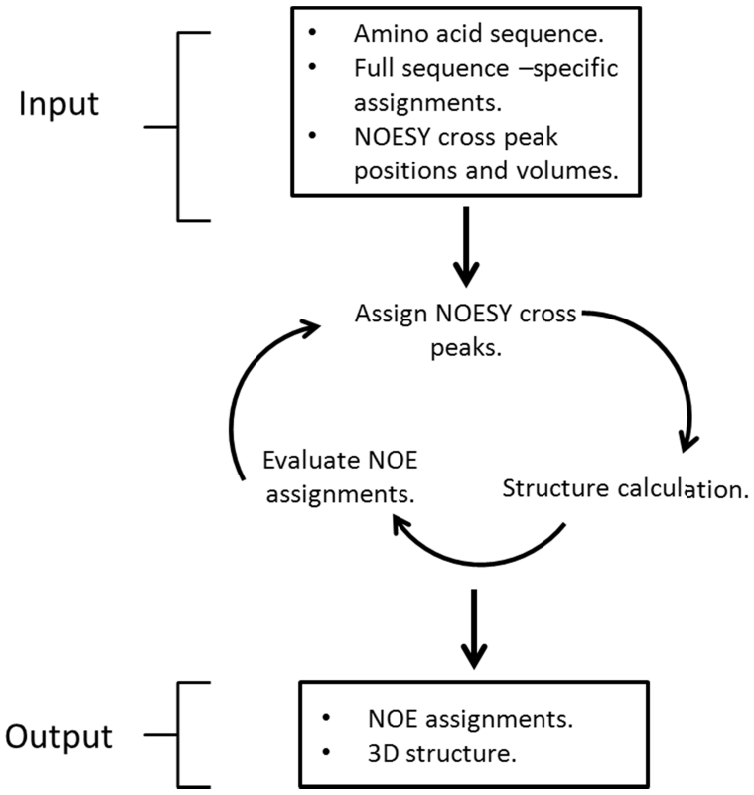

Figure 5.4-3: Automatic NOESY assignment and structure calculation using CYANA. A schematic of the automated structure calculation used by the programme CYANA is shown. The programme utilises an iterative procedure to assign the NOE peaks and calculate the protein structure.

5.4.3 ^{13}C NOE Distance Constraints.

The ^{13}C-NOESY-HSQC is analogous to the ^{15}N-NOESY-HSQC described above, with the exception that the ^{15}N dimension is replaced by a ^{13}C dimension. The ^{1}H-^{13}C-HSQC (Figure 5.4-4) spectrum contains a single peak for each proton bound to a carbon and so is more crowded than the ^{1}H-^{15}N HSQC spectrum. The ^{13}C-NOESY-HSQC extends the ^{1}H-^{13}C-HSQC in to a third ^{1}H-NOESY dimension. The result is a NOESY 'strip' for each ^{1}H-^{13}C-HSQC peak allowing the determination of distance restraints between any carbon-bound proton and all other protons. This spectrum provides a great deal more structural information than the ^{15}N-NOESY-HSQC as distance constraints between side chain nuclei are obtained.

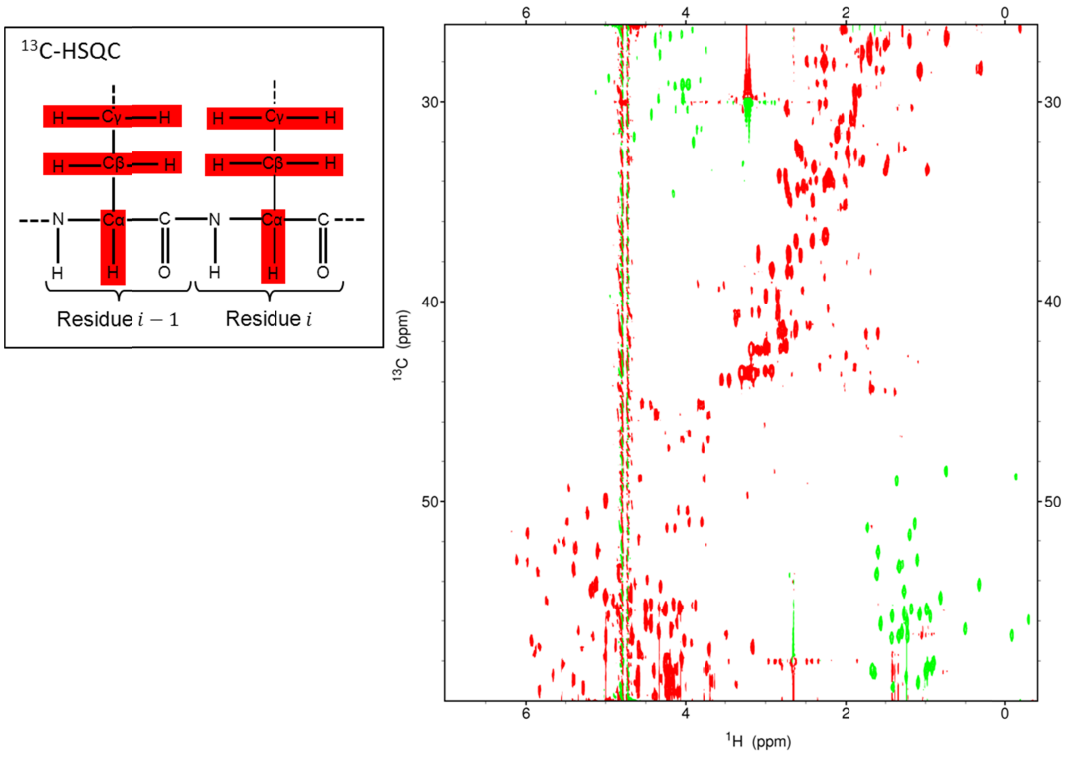

Figure 5.4-4: ^{1}H-^{13}C-HSQC spectrum of DC-SIGNR CRD. The ^{13}C-HSQC contains one peak for every proton bound to a carbon in the protein. For this reason it is more crowded than the ^{15}N-HSQC spectrum. Negative peaks are shown in green, this is due to peaks that are outside of the spectral window being aliased.

Similarly to the ^{15}N-NOESY-HSQC, several mixing times were investigated in order to determine the correct mixing time for the ^{13}C-NOESY-HSQC. Figure 5.4-5 shows ^{1}H-^{1}H planes (all signals in the 3D experiment are collapsed on to a single plane) of the ^{13}C-NOESY-HSQC experiment at different mixing times. A substantial increase in the number of NOE signals was observed with a mixing time of 150 ms, and so this was used to acquire a full 3D dataset. Mixing times greater than 150 ms were not investigated in order to avoid contributions from spin-diffusion. The number of peaks present in Figure 5.2-1 at 150 ms suggests that a sufficient number of additional distance constraints could be obtained from the ^{13}C-NOESY-HSQC which could improve the structure calculation. Work is on-going to fully analyse the ^{13}C-NOESY-HSQC and perform a structure calculation including these distance restraints. If the additional distance restraints produce a converged structure further structural refinement will be required.

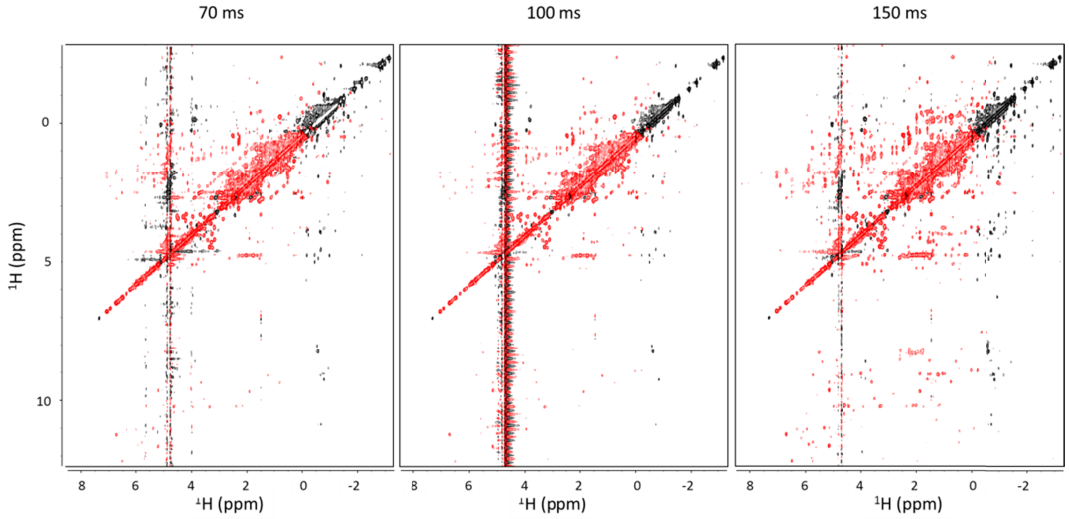

Figure 5.4-5: ^{13}C-NOESY-HSQC planes at varying mixing times. ^{13}C-^{1}H planes of the ^{13}C-NOESY-HSQC are shown (the third proton dimension is collapsed in to a single plane so all peaks are shown simultaneously). Very few NOE peaks are present at 70 and 100 ms mixing times. A mixing time of 150 ms results in the greatest number of peaks and so this will be used for structure calculations in the future.

5.5 Conclusion.

The backbone assignment along with the H_α assignments obtained from the H(CCO)NH spectrum were used to predict the secondary structure of the DC-SIGNR CRD using the programme TALOS+ which is based on the chemical shift index method. The secondary structure prediction is in good agreement with the crystal structure (1K9J) with the exception of the C-terminal region which lacks predictions due to the presence of un-assigned residues. The size of the CRD prevented assignment of the full side-chain (using the TOCSY-HSQC experiment) due to fast ^{15}N backbone relaxation causing loss of signal during the TOCSY mixing time. Instead, HCCH-TOCSY and CCH-TOCSY experiments were acquired enabling the first reported side-chain assignment of the DC-SIGNR CRD (Appendix A). The assignment paves the way for a full solution structure determination of the CRD.

While the ^{15}N-NOESY-HSQC spectrum (with a 200 ms mixing time) provides over 600 NOE cross peaks which could produce valid distance restraints, simultaneous structure calculation and NOE assignment using the programme CYANA did not converge using these NOE peaks alone. As a result, the quality of the ^{13}C-NOESY-HSQC spectrum was investigated because the ^{13}C-^{1}H NOE correlations provide more structural information. A mixing time of 150 ms produces many NOE cross peaks with sufficient signal/noise in the ^{13}C-NOESY-HSQC spectrum. Work is on-going to determine if the addition of these restraints could help the solution structure to converge.

6 Calcium Binding and pH Dependence.

6.1 Introduction.

This chapter investigates the calcium binding characteristics of the DC-SIGNR CRD, incorporating the presence of EDTA (a calcium chelating agent) and low pH conditions in an effort to understand further the DC-SIGNR-calcium interactions which, as yet, have not been investigated in detail.

EDTA can be used to probe the presence of calcium in solution via NMR as the 1D ^{1}H spectra of free and calcium-bound EDTA can be easily distinguished (Han & Ba, 2004; Han, 2007). EDTA binds strongly to calcium and should compete with the DC-SIGNR CRD for calcium ions. As a result, addition of EDTA to the DC-SIGNR CRD samples was used to confirm that the CRD was in the apo state because the presence of free EDTA NMR signals confirmed the absence of calcium. If EDTA has a higher affinity for calcium than the CRD, the addition of excess EDTA to the holo CRD (calcium-bound) should result in chelation of all calcium ions leaving the CRD in the apo form (which can be confirmed by comparing the HSQC spectrum with that of a known apo sample).

In addition, the effect of pH on both calcium and glycan binding to the CRD was investigated. This is of interest as it has been suggested that DC-SIGNR acts as an endocytic receptor by internalising pathogens and releasing them in the low pH environment of cytosolic compartments (Gramberg et al, 2008; Guo et al, 2004; Snyder et al, 2005b). It has been confirmed that many other C-type lectins have this function, including DC-SIGN (Engering et al, 2002). However, the fact that DC-SIGNR lacks a YxxL motif present in DC-SIGN, which is known to act as an internalisation signal, brings this into debate. In addition, it has been reported that DC-SIGNR does not internalise ligand and has no pH dependence when transfected into rat fibroblasts (Guo et al, 2004). Any knowledge of the pH dependence (and by extension the endocytic activity) of the CRD will provide valuable insight in to the potential biological function of DC-SIGNR. The pH dependence of the DC-SIGNR CRD was investigated by acquiring HSQC spectra of the holo and glycan-bound CRD at pH 6.8 and pH 4.2. Any changes to the CRD as a result of decreasing pH, either conformational changes or release of calcium/glycan, should be apparent in the HSQC spectra as these changes will result in chemical shift perturbations. We have already seen in §4.4.1 that there are substantial differences between the HSQC spectra of the apo and holo CRD and so it is reasonable to expect that decreasing the pH will result in significant changes in the HSQC spectrum if calcium is released.

6.2 EDTA as a Probe for Calcium Binding.

The ^{1}H chemical shifts of EDTA in the free and calcium bound state have been reported previously (Han & Ba, 2004; Han, 2007). The free and calcium-bound EDTA chemical shifts are easily distinguishable: free EDTA results in two sharp singlet peaks due to the acetate-methylene (labelled 1 in Figure 6.2-1) and amine-methylene protons (labelled 2 in Figure 6.2-1). Calcium binding causes splitting of the acetate-methylene proton peak while a singlet remains for the amine-methylene protons. In the calcium bound form, both peaks are shifted up-field with respect to the free EDTA chemical shifts. Consequently, the addition of EDTA can be used to screen for the presence of calcium in a solution (EDTA has been used to determine the quantity of calcium, as well as other metal ions in hard water (Yappert & DuPre, 1997)). EDTA is expected to compete with the DC-SIGNR CRD for calcium and so the differences in the ^{1}H spectrum of free and calcium bound EDTA can be used to determine whether the CRD is in the apo (calcium free) or holo (calcium bound) state. If the addition of EDTA to the CRD results in two singlet peaks (free EDTA peaks) then we can conclude that the CRD is in the apo form because no calcium is present. If EDTA has a higher affinity for calcium than the CRD then addition of EDTA to a holo CRD sample should result in complete removal of calcium, leaving the CRD in the apo form.

Figure 6.2-1 (A) shows the aliphatic region of the ^{1}H spectrum of apo DC-SIGNR containing EDTA. Only signals arising from free-EDTA are present confirming the complete absence of calcium and hence the CRD must be in the apo state. This confirms that the HSQC spectrum in §4.4, Figure 4.4-1 is indeed of

the apo CRD. Spectrum (B) in Figure 6.2-1 shows the ^{1}H spectrum of a holo DC-SIGNR CRD sample (containing 4 mM $CaCl_2$) following addition of 1 mM EDTA. Only signals from calcium bound EDTA are present. The protein concentration is 0.7 mM so the addition of 4 mM $CaCl_2$ ensures that calcium is in excess and the CRD is in the holo form. As a result, free calcium is present in solution and binds to the EDTA producing the peaks in Figure 6.2-1 (B). The addition of excess EDTA (60 mM) to the holo CRD results in chelation of all free calcium and the presence of free EDTA in solution resulting in the spectrum in Figure 6.2-1 (C).

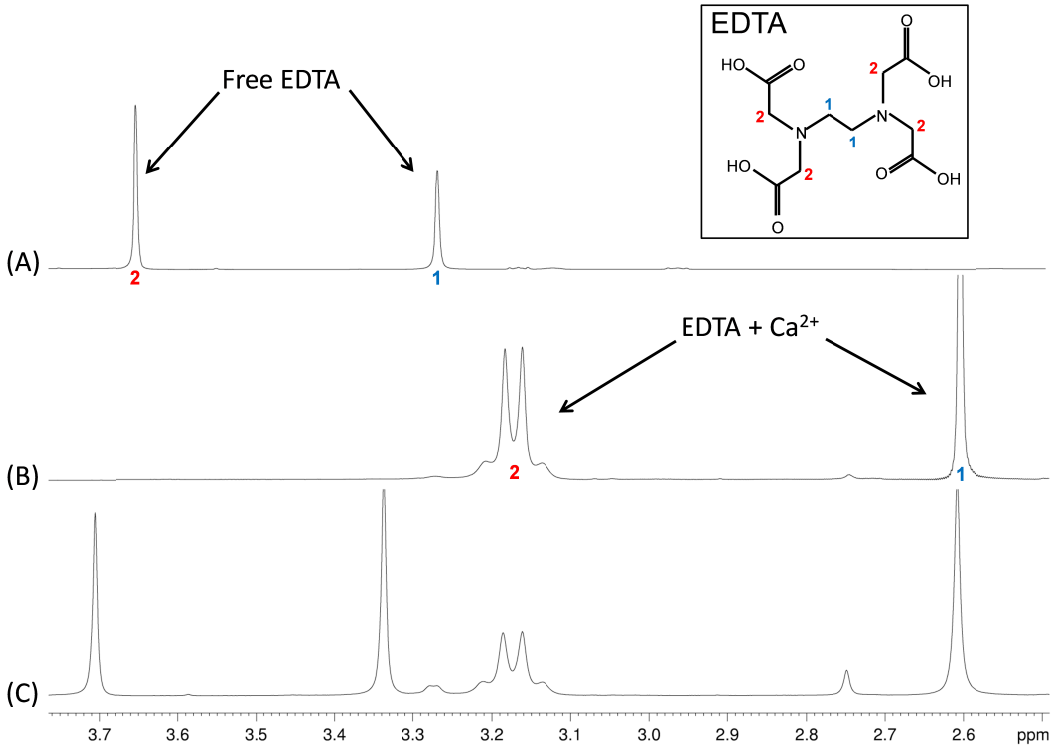

Figure 6.2-1: 1D-^{1}H spectra of EDTA added to (A) apo DC-SIGNR, (B) holo DC-SIGNR and (C) excess EDTA added to holo DC-SIGNR. All spectra were recorded at pH 6.8. EDTA can be used to probe the presence of calcium as the 1D spectrum of free and calcium-bound EDTA are easily identified. Assignments are mapped on to the EDTA structure and the pH of all spectra was kept constant. (A) The addition of EDTA to apo DC-SIGNR CRD produces only free EDTA peaks confirming that there is no calcium present and that the protein is indeed in the apo state. In contrast, addition of EDTA to holo DC-SIGNR (B) produces only calcium-bound EDTA peaks as the free calcium interacts with EDTA. Adding excess EDTA to the holo sample (C) results in the appearance of additional free EDTA peaks.

Interestingly, there is evidence of calcium exchange between free and calcium bound EDTA. Figure 6.2-2 shows the effect on the ^{1}H spectrum of holo DC-SIGNR as the EDTA concentration is increased. Initially, the calcium bound EDTA peaks increase in intensity until all of the free calcium has been chelated and free EDTA peaks begin to emerge. However, while the free EDTA peaks increase in intensity, the bound EDTA peaks begin to decrease and broaden suggesting exchange between the free and bound EDTA (calcium is transferring between EDTA molecules). This is possibly due to sub-optimal chelation by EDTA at pH 6.8 (EDTA chelates metal ions most efficiently at pH ≥ 10 where the fully deprotonated form is dominant) (Skoog et al, 1994). However, the quality of NMR spectra suffers at higher pH values (due to the increased rate of proton exchange causing line broadening) and the study of DC-SIGNR at close to physiological conditions is desirable, therefore EDTA titrations at higher pH values were not acquired.

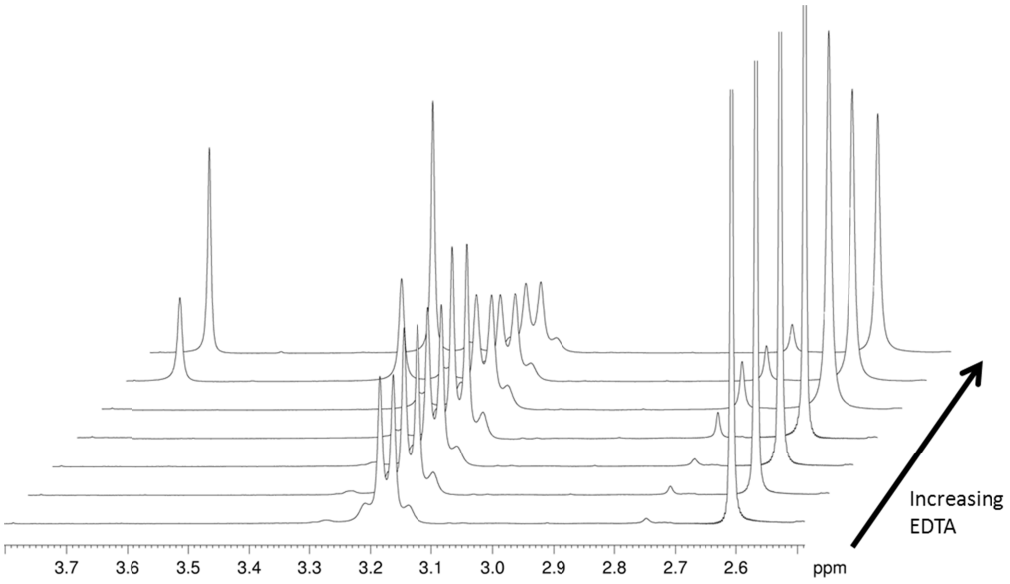

Figure 6.2-2: ^{1}H stack plots of EDTA titrations of holo DC-SIGNR. As the amount of EDTA added to holo DC-SIGNR is increased the calcium-bound EDTA peaks increase in intensity due to free calcium binding to the EDTA. When all of the free calcium is bound to EDTA, free-EDTA peaks appear. Interestingly, the calcium-bound EDTA peaks also decrease at this point suggesting exchange between the free and calcium-bound EDTA. The pH was kept constant at 6.8 throughout the titration.

The fact that the HSQC spectrum of holo DC-SIGNR containing excess (60 mM) EDTA is not the same as the spectrum of apo DC-SIGNR (Figure 6.2-3) also suggests that exchange between EDTA and calcium may be taking place. The presence of free EDTA peaks in the ^{1}H spectrum of the holo DC-SIGNR sample containing 60 mM EDTA implies that all of the calcium is sequestered (if we assume that EDTA binds calcium with higher affinity than the CRD) and so there should be no calcium remaining which is capable of binding to the CRD. Therefore, the differences between the HSQC spectra suggest that there is exchange between the calcium bound to DC-SIGNR and free EDTA. This results in the chemical shifts in the HSQC spectrum appearing at a position between the apo, holo (three calcium ions bound), and possibly 1 and/or 2 calcium ion-bound forms. In fact, the chemical shifts of the excess EDTA sample are not a weighted

average between the apo and holo forms implying that there are other species present (most likely the 1 and/or 2 calcium ion-bound forms). The variable intensity of the spectrum could be further evidence of exchange between these multiple species. Thus, the HSQC spectrum of the CRD containing 4 mM $CaCl_2$ and 60 mM EDTA (Figure 6.2-3, green) represents a calcium depleted form of the CRD. This means that regions of the DC-SIGNR CRD can bind calcium with very high affinity, potentially on the same order as the affinity of EDTA for calcium (determined previously as ~0.055 mM (Brinley et al, 1975; Christensen et al, 2003)). However, this is significantly stronger than dissociation constants published for other C-type lectins such as the asialoglycoprotein receptor, which exhibits an apparent K_d of ~330 µM for calcium at calcium-binding sites 1 and 2 and ~19 mM for calcium-binding site 3 at pH 6.9 (Onizuka et al, 2012). Unfortunately, this exchange coupled with difficulties associated with the removal of calcium (due to the apparent high affinity of the CRD for calcium) means that the apo form cannot be retrieved from the holo CRD. However, the HSQC spectra of holo DC-SIGNR both in the absence and presence of EDTA, but with addition of excess calcium, are the same (Figure 6.2-4 (B)).

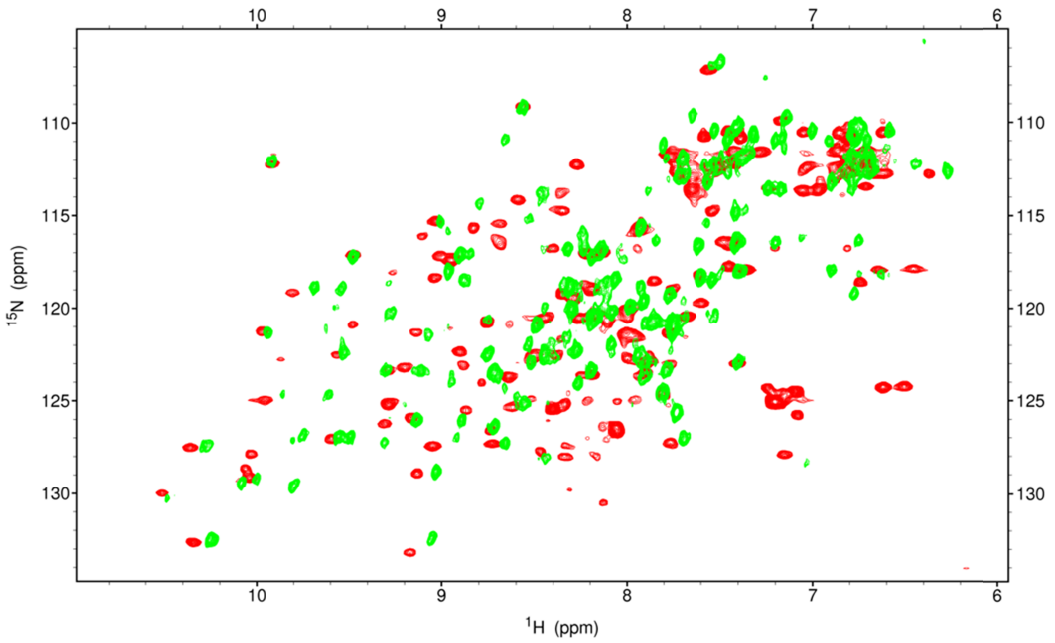

Figure 6.2-3: Overlay of HSQC spectra of apo DC-SIGNR (red) and holo DC-SIGNR with excess EDTA (green) both at pH 6.8. Despite the addition of 60 mM EDTA to the holo DC-SIGNR CRD sample the HSQC does not resemble the apo HSQC as expected. This suggests that at this pH the chelating ability of EDTA is not sufficient to remove all calcium from the CRD. This highlights that the calcium affinity of DC-SIGNR is high.

Figure 6.2-4 shows the effect of increasing calcium on holo DC-SIGNR CRD in the presence of a fixed amount of EDTA. Initially, free EDTA is present in solution as seen by the free EDTA peaks present in the 1D spectrum (Figure 6.2-4 (A)). As the calcium concentration increases all of the free EDTA becomes bound resulting in loss of the free EDTA peaks and increased intensity of the calcium bound peaks. Under these conditions (60 mM EDTA, 70 mM $CaCl_2$) the HSQC spectrum overlays almost exactly with the spectrum of holo DC-SIGNR (Figure 6.2-4 (B)) suggesting that calcium exchange between protein and bound EDTA is negligible at these concentrations. This means that EDTA has no effect on the holo HSQC spectrum but detailed information on DC-SIGNR-calcium binding cannot be obtained in the presence of EDTA due to calcium exchange.

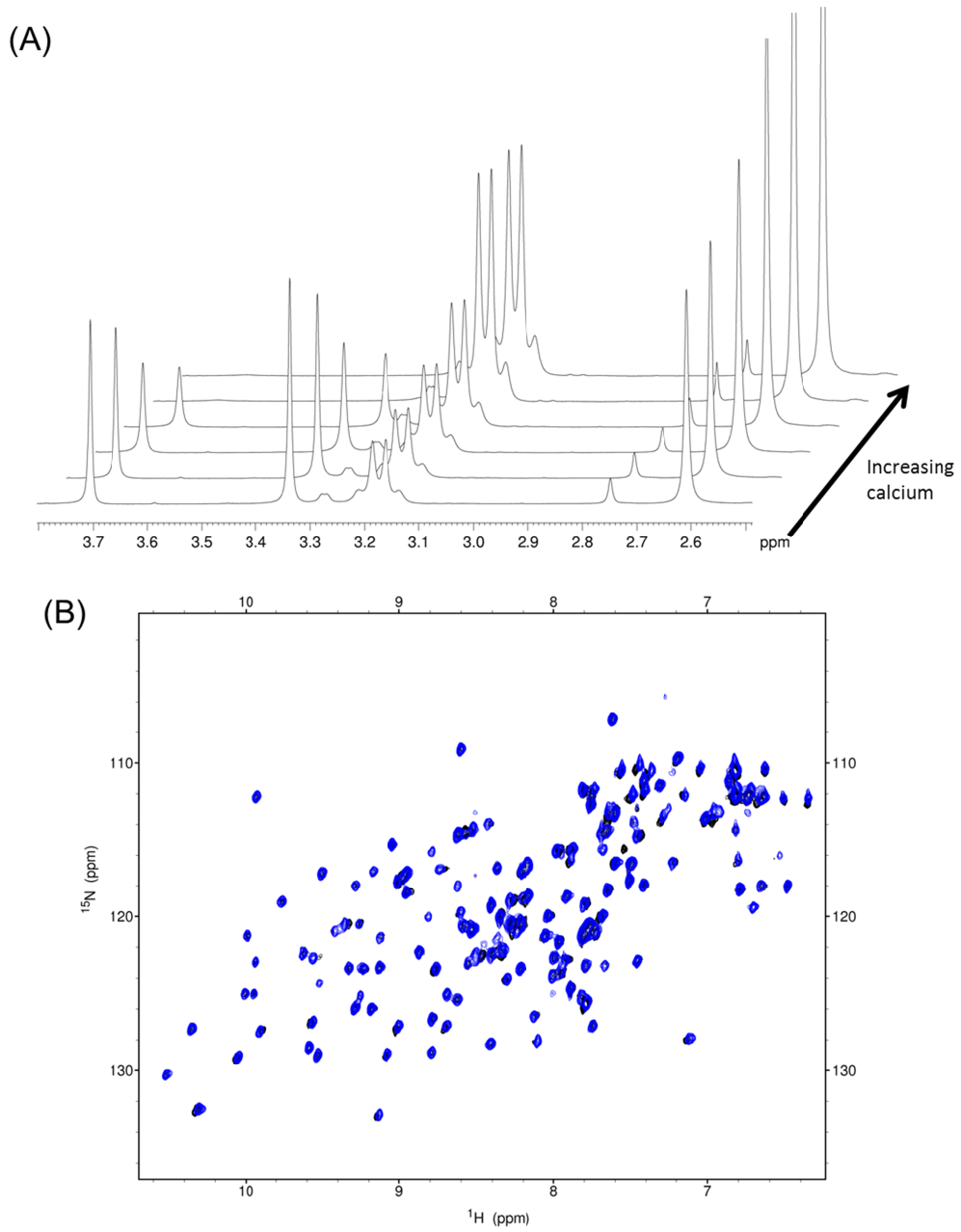

Figure 6.2-4: (A) ^{1}H stack plots of calcium titrations on holo DC-SIGNR in the presence of EDTA and (B) overlay of HSQC spectra of holo DC-SIGNR in absence of EDTA (blue) and presence of both excess EDTA and calcium (black) at pH 6.8. (A) Titration of calcium into the holo CRD containing excess EDTA causes the free EDTA peaks to disappear and the calcium-bound peaks to increase in intensity. The pH was kept constant at 6.8 throughout the titration. (B) The HSQC spectra of holo DC-SIGNR in absence of EDTA (blue) and following addition of 60 mM EDTA and 70 mM $CaCl_2$ (black) are extremely similar. This suggests that there is sufficient calcium to saturate both the EDTA and the CRD.

6.3 The pH Dependence of DC-SIGNR Glycan Release.

While the complete physiological role of DC-SIGNR is not known with confidence it is known to have roles in cell-adhesion (Yabe et al, 2010), cell signalling (Johnson et al, 2012) and may be involved in initiating primary immune response through binding to ICAM-3 (Bashirova et al, 2001). There is also speculation that DC-SIGNR could act as an endocytic receptor by internalising pathogens. Other C-type lectins have endocytic functionality (Engering et al, 2002; Valladeau et al, 2000; Wragg & Drickamer, 1999) and there are reports that DC-SIGNR can internalise glycoprotein ligands (Gramberg et al, 2008). Endocytic C-type lectins are believed to release glycoproteins/pathogens in the endosome as a result of calcium release in the low pH environment (Mellman et al, 1986). However, there is a report which shows that DC-SIGNR-ligand binding is not pH dependent and that DC-SIGNR does not internalise glycoproteins (Guo et al, 2004). In addition, DC-SIGNR lacks an internalisation signal (a YxxL motif in the cytoplasmic domain) that is present in the unequivocally endocytic receptor DC-SIGN (Soilleux et al, 2000). As a result, some authors argue that DC-SIGNR has evolved to perform different (as yet uncharacterised) biological functions than DC-SIGN that do not necessarily involve ligand internalisation. This hypothesis was investigated utilising NMR spectroscopy. A comparison of HSQC spectra at varying pH values could provide information on the conformational state of the CRD and shed light on the pH dependence of DC-SIGNR ligand binding.

6.3.1 Decreasing the pH Causes Significant Changes in the HSQC Spectrum of Holo DC-SIGNR CRD.

Figure 6.3-1 compares HSQC spectra of holo DC-SIGNR CRD at pH 6.8 (dark blue) and 4.2 (light blue). It is clear that large conformational/dynamics changes occur throughout the CRD as a result of decreased pH as many chemical shifts are perturbed. The most likely explanation for this is the release of calcium. However, the holo HSQC spectrum at pH 4.2 does not overlay exactly with the HSQC spectrum of apo CRD at pH 6.8 (Figure 6.3-2) indicating that decreasing the pH contributes to an additional conformational or dynamics change or that there is exchange between multiple species (apo, holo and calcium depleted) of the CRD. These pH induced changes to the holo-CRD HSQC spectrum are reversible since increasing the pH back to 6.8 returns the spectrum to the holo form, indicating that the CRD regains the ability to bind ligand when it returns to the higher pH extracellular environment.

Although the spectra of the holo CRD at low pH (light blue) and apo CRD (red) (Figure 6.3-2) do not overlay exactly, there are patterns of peaks that appear at low pH that are only present in the apo spectrum. Most notably the peaks circled in Figure 6.3-2 appear as the pH is decreased, these peaks are only ever present in the apo spectrum. Unfortunately, these peaks are not assigned (as they are not present in the holo HSQC spectrum) but they are possibly aliased arginine side-chain peaks. The significant changes in the HSQC spectrum (Figure 6.3-1) along with the appearance of the peaks highlighted in Figure 6.3-2 lead us to speculate that calcium is released at low pH.

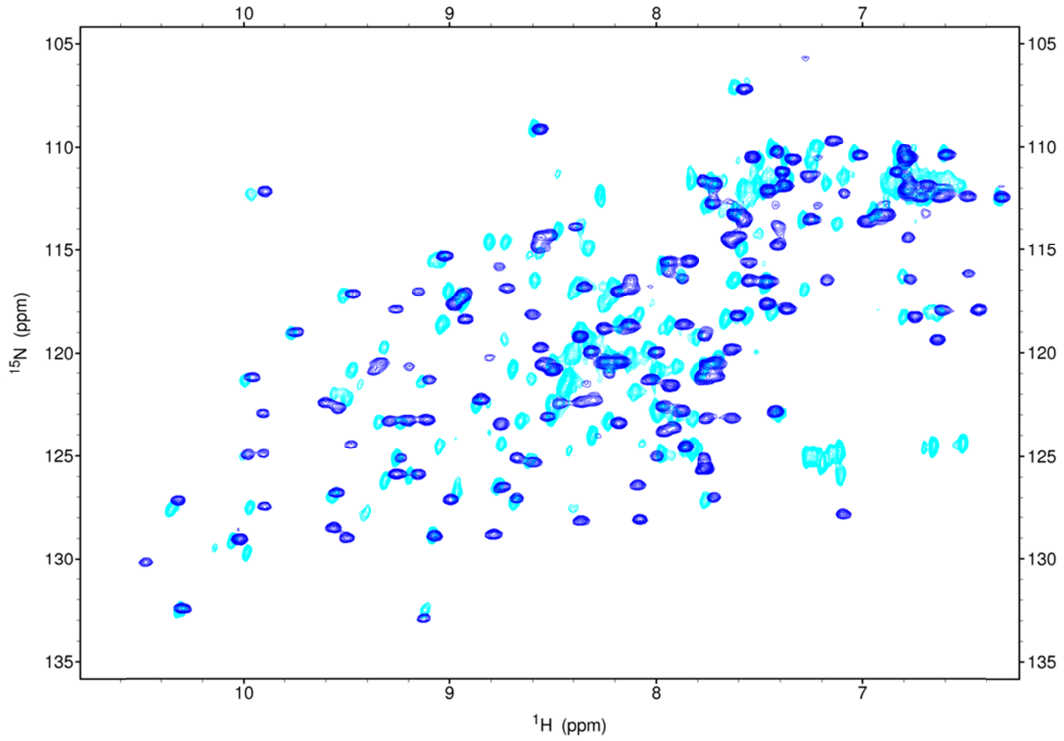

Figure 6.3-1: HSQC spectra of holo DC-SIGNR at pH 6.8 (dark blue) and 4.2 (light blue). Decreasing the pH of the holo DC-SIGNR CRD sample causes chemical shift perturbations and variations in intensity throughout the HSQC indicating that significant conformation/dynamics changes occur. The most likely explanation is that calcium is released at low pH. However, the spectrum of the holo CRD at pH 4.2 is not the same as the apo CRD at pH 6.8 indicating that the pH may reversibly affect the protein conformation.

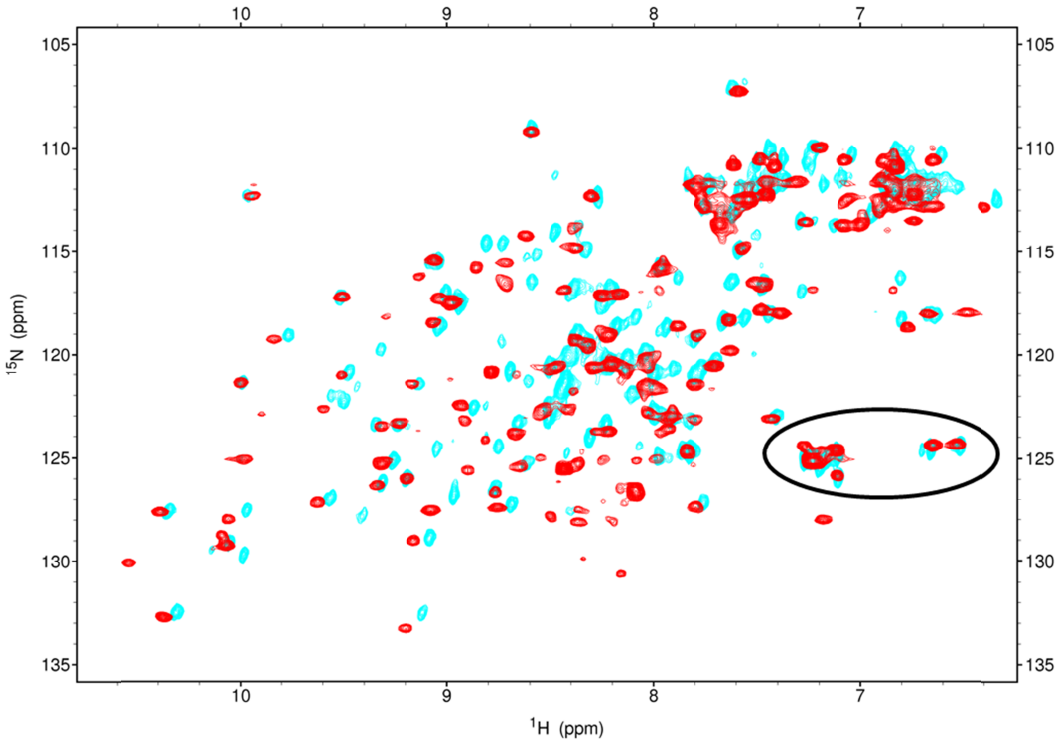

Figure 6.3-2: HSQC spectra of holo DC-SIGNR at pH 4.2 (light blue) and apo DC-SIGNR at pH 6.8 (red). Although the HSQC spectra of holo DC-SIGNR at low pH and apo at neutral pH are similar, which could mean that calcium is released at low pH, there are some chemical shift differences indicating that the low pH contributes to an additional conformation or dynamics change. Some peaks that are only present in the apo and low pH spectra are circled.

6.3.2 *Evidence of Glycan Release at Low pH.*

The large number of chemical shift perturbations in the HSQC spectra of holo DC-SIGNR at low pH compared to pH 6.8 suggests the possibility that calcium release coupled with a pH induced conformational/dynamics change in the DC-SIGNR CRD could be responsible for ligand release in the cell cytosol. The effect of pH on ligand binding was tested by decreasing the pH of holo DC-SIGNR in the presence of 10 mM $GlcNAc_3$. While this ligand does not contain mannose, GlcNAc residues are found at the reducing end of high-mannose structures and GlcNAc has been implicated in interactions between DC-SIGNR and West Nile virus (Davis

et al, 2006). Binding of $GlcNAc_3$ to DC-SIGNR was confirmed via chemical shift perturbations observed for a number of HSQC signals (Figure 6.3-3). Decreasing the pH of the $GlcNAc_3$-bound DC-SIGNR CRD to 4.2 causes significant chemical shift perturbations (more perturbations than induced by glycan binding) suggesting a drastic, and global change in the CRD as a result of decreasing the pH. Figure 6.3-4 compares the HSQC spectra of holo DC-SIGNR (no glycan present, light blue) and $GlcNAc_3$-bound DC-SIGNR (purple) both at pH 4.2. The similarity between the two spectracould be a direct result of ligand (and calcium) release. The peaks circled in Figure 6.3-2 and Figure 6.3-3 are due to aliased arginine side-chain amides. These peaks are visible in the low pH spectra of both holo and glycan-bound DC-SIGNR CRD but at pH 6.8 are not visible due to broadening / chemical shift changes. More work is required to determine the cause of these changes. Interestingly these peaks are also visible in the apo spectrum of DC-SIGNR CRD at pH 6.8 (Figure 6.3-2, red). There are differences between the two spectra in Figure 6.3-2 which could indicate that reducing the pH does not completely abolish ligand/calcium binding resulting in chemical shifts intermediate between bound and un-bound states. This could also explain the discrepancies between the HSQC spectra of holo CRD at pH 4.2 and apo at pH 6.8 (Figure 6.3-2).

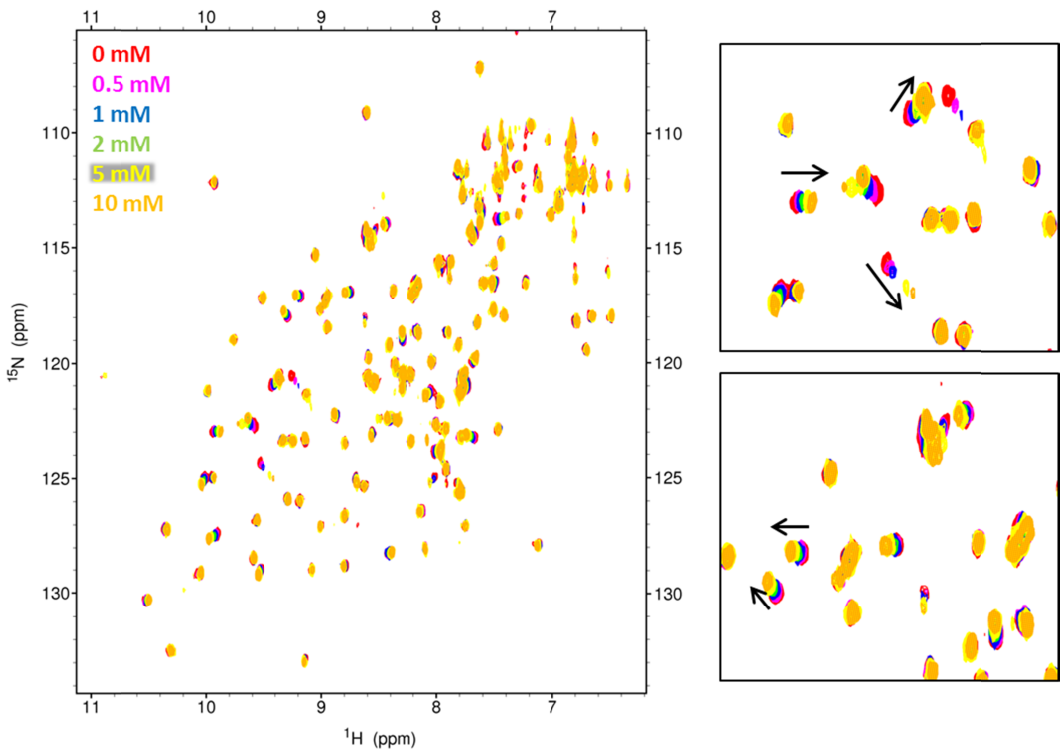

Figure 6.3-3: HSQC spectra of $GlcNAc_3$ titration of holo DC-SIGNR. Chemical shift perturbations confirm that there is an interaction between the DC-SIGNR CRD and $GlcNAc_3$.

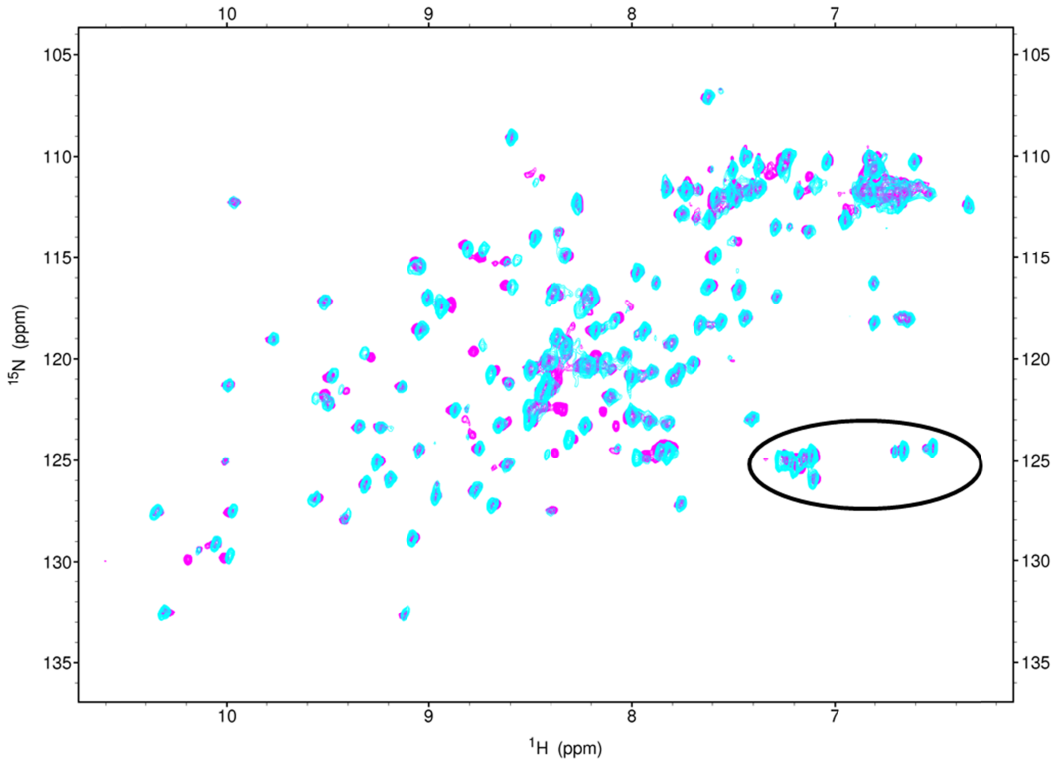

Figure 6.3-4: HSQC spectra of DC-SIGNR CRD at pH 4.2 in holo (light blue) and glycan bound form (purple). Dropping the pH of the DC-SIGNR CRD in holo form and in the presence of glycan has a similar effect on the HSQC spectrum. However, there are a number of significant chemical shift perturbations which may indicate that glycan binding is not completely abolished at low pH. Peaks which appear at low pH that are only present in the apo HSQC spectrum are circled.

Figure 6.3-5 highlights that the effect of pH on the glycan-bound CRD is complex. This is expected as there are potentially multiple un-binding events (calcium and glycan) along with conformational and dynamics changes associated with the change in pH. There is evidence of line broadening (circled in Figure 6.3-5) which is indicative of intermediate exchange on the NMR timescale and could indicate that ligand/calcium binding has not been completely abolished. In addition, linear shifts represent fast exchange on the chemical shift timescale and could be a direct result of ligand or calcium release (fast and intermediate chemical shift perturbations are discussed in more detail in §7.1.1). Interestingly, some chemical shifts follow a curved perturbation path which indicates that more than one molecule is involved, suggesting that both ligand and calcium are released at low pH.

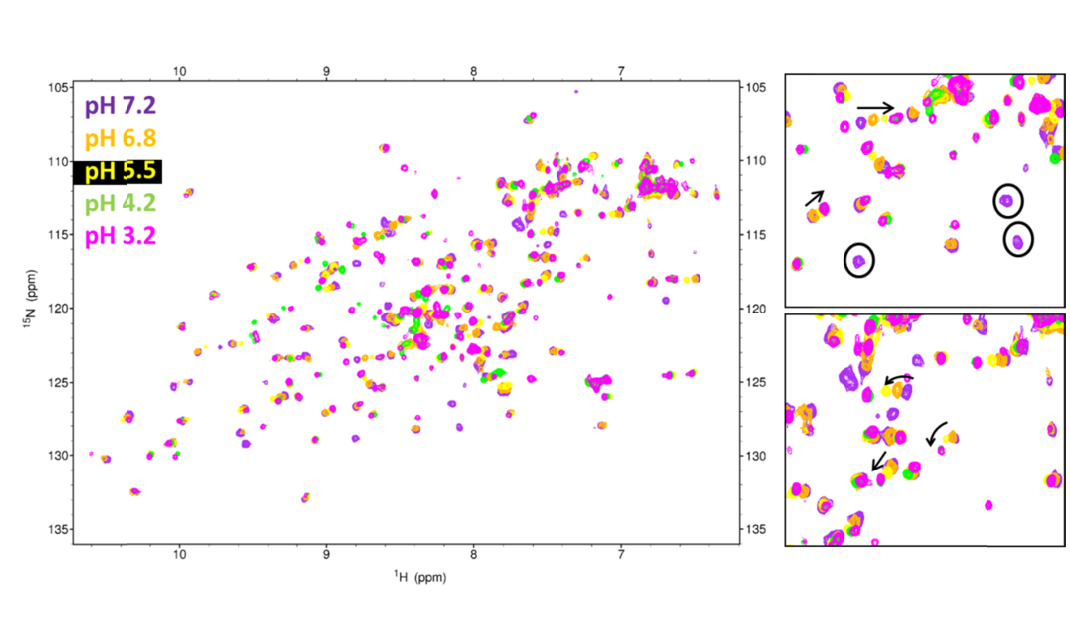

Figure 6.3-5: HSQC spectra of pH titration of DC-SIGNR containing 10 mM $GlcNAc_3$. The effect of varying pH on glycan bound DC-SIGNR was tested. Increasing the pH to 7.2 causes little to no chemical shift perturbation (compared to pH 6.8) while decreasing the pH causes significant perturbations throughout the protein. Some peaks are perturbed linearly while others have a curved path which is indicative of multiple molecules binding/unbinding. In addition, some peaks broaden as the pH is lowered which indicates intermediate exchange on the NMR chemical shift time-scale.

In order to confirm that DC-SIGNR releases ligand at low pH protein was bound to a 2 mL mannose-Sepharose column (as described in the Materials and Methods §2.7.1) and eluted using an EDTA-free buffer at pH 4.2. Figure 6.3-6 confirms that DC-SIGNR dissociates from the mannosylated resin, and elutes from the column when the pH is decreased which supports the NMR data above. The protein is predominantly found in the third 1 mL elution fraction at low pH, which is in contrast to the EDTA elution method wherein the majority of protein is consistently eluted in the second 1 mL fraction. In addition, using the low pH method results in a small amount of protein present in the fourth 1 mL elution fraction while the EDTA method results in all protein eluting by the third 1 mL fraction. This indicates that lowering the pH does not completely abolish ligand binding. This supports the NMR data suggesting that the differences between the HSQC spectra of holo CRD and glycan bound CRD at pH 4.2 are due to exchange between bound and un-bound states as a result of decreased, but not abolished, ligand binding. Carbohydrates are weak acids with pKa values ranging from 12-14 (Lange & Speight, 2005) and so a pH drop from 6.8 to 4.2 will not have any substantial effect on the glycan protonation state, confirming that the changes in the HSQC are due to the pH dependence of the CRD.

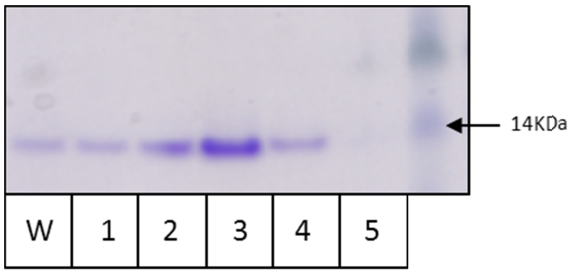

Figure 6.3-6: Eluting from mannose-Sepharose column with low pH. The mannose-Sepharose column was eluted using 20 mM hepes, 150 mM NaCl pH 4.2. The majority of protein elutes in fraction 3 confirming the DC-SIGNR is capable of releasing ligand at low pH. In contrast to the EDTA elution method, there appears to be a small amount of protein in fraction 4 suggesting that lowering pH does not abolish binding as efficiently as EDTA addition.

6.4 Conclusion.

This chapter describes changes in the HSQC spectrum of the DC-SIGNR CRD in the presence and absence of calcium, upon addition of EDTA, addition of glycan, and at low pH. Several HSQC spectra were acquired which represent the CRD in different forms: apo, holo, calcium-depleted, glycan-bound, and low pH. These are summarised schematically in Figure 6.4-1. This study has provided insight in to the calcium binding characteristics of the CRD, which has not been studied in detail previously. The data suggest that the CRD binds calcium with high affinity (once added in excess, Ca^{2+} cannot be completely removed from the CRD using EDTA alone) and that calcium and ligand are released at low pH supporting the argument that DC-SIGNR has a role as an endocyctic receptor (Gramberg et al, 2008; Snyder et al, 2005b).

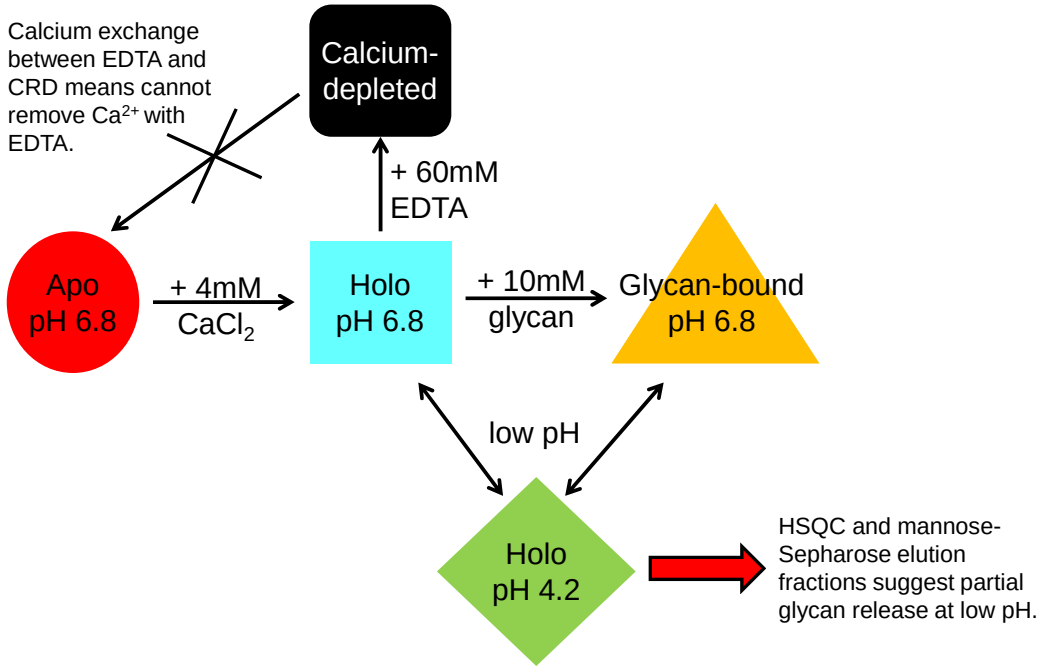

Figure 6.4-1: Schematic representation of the data described in this chapter and the CRD species that have been observed. Each shape represents the HSQC of the CRD in different conditions. Under different conditions the chemical shifts in the HSQC spectra change. However, the HSQC of the holo CRD and the glycan-bound CRD at low pH are extremely similar. This, coupled with the low pH mannose-Sepharose elutions suggests that the CRD releases glycan at low pH.

EDTA provided an effective means of determining the presence of excess free calcium in protein samples. The acquisition of ^{1}H spectra of the apo CRD following addition of EDTA confirms that no calcium is present in the sample and thus the protein is in the apo state. Equally, addition of EDTA to the holo CRD samples confirms the presence of excess calcium. Attempts to acquire a HSQC spectrum of apo DC-SIGNR in the presence of calcium by addition of excess EDTA were unsuccessful. This is due to calcium exchange between the CRD and free EDTA resulting in chemical shifts at positions intermediate between the apo and holo spectra producing a HSQC spectrum of the calcium-depleted CRD. This indicates that DC-SIGNR binds calcium with high affinity and that the calcium binding sites have different affinities. This problem might be resolvable by increasing the pH to >9.0 and thus increasing the affinity of EDTA for calcium.

However, this would have a detrimental effect on the spectral signal/noise and so was not explored. The use of chelating agents with stronger calcium affinity, such as EGTA, could be used as an alternative.

The effect of pH on calcium and ligand binding was investigated as there are conflicting reports in the literature and because any insight in to mechanisms of ligand release could help in understanding the biological role of DC-SIGNR. Decreasing the pH of the holo CRD causes substantial changes to the HSQC chemical shifts which indicate structural rearrangement that could be a result of calcium release. However, the HSQC spectrum of holo CRD at low pH is not the same as that of the apo CRD indicating that additional conformational/dynamics changes occur as a result of lowering the pH or that not all Ca^{2+} is released at low pH. Similarly, the HSQC spectra of DC-SIGNR CRD bound to $GlcNAc_3$ undergoes extensive chemical shift perturbations at low pH. A pH titration highlights the chemical shift perturbations and confirms the complexity of the effect of pH on the CRD. This could be a result of ligand and calcium release causing curved perturbations as well as line broadening. This line broadening, along with the fact that the low pH HSQC spectra do not overlay with the apo HSQC spectrum, indicates that decreasing the pH does not completely abolish ligand binding. Protein was successfully eluted from the mannose-Sepharose column using pH 4.2 buffer confirming that DC-SIGNR-glycan interactions are pH dependent. However, protein was observed in elution fraction 4 (this is not observed when using the EDTA elution method) suggesting that some degree of mannose binding is maintained at low pH. This supports previous work which shows that, while DC-

SIGNR releases ligand at low pH, the effect of pH on DC-SIGNR is not as substantial as the effect on DC-SIGN (Gramberg et al, 2008).

While these results illuminate some interesting effects of pH on the glycan binding ability of the DC-SIGNR CRD, they are not conclusive and further work is required to achieve a deeper insight in to the mechanism of ligand release at low pH. On-going work includes investigation of the apo CRD at low pH (the difficulties in obtaining calcium free protein prevented analysis of this within the timescale of this project), investigation of the effect of pH on mannose-containing ligands (particularly branched, disease-associated ligands), and the determination of protein-ligand dissociation constants at low pH.

7 Glycan Binding.

7.1 Introduction.

The backbone assignment of the holo DC-SIGNR CRD (described in §4.3) provides a 'map' of the protein and allows the effects of ligand binding on the DC-SIGNR CRD to be investigated directly. The backbone chemical shifts are sensitive to changes in the chemical environment of a given nucleus, and so direct interaction with a ligand along with ligand induced conformational or dynamics changes will result in the perturbation of chemical shifts in the HSQC spectrum. This allows for the easy identification of residues/regions of the CRD that are affected by ligand binding. This method of probing protein-ligand interactions allows rapid investigation of a range of ligands in solution eliminating the problems which arise from difficult-to-crystallise ligands. Due to their inherent flexibility (Woods et al, 1998), branched oligosaccharides can be a challenge to crystallise and crystal structures of the DC-SIGNR CRD in complex with large, disease-associated ligands remain elusive. In particular, utilising solution state NMR has allowed the investigation of $Man_9GlcNAc$ derived from the HIV envelope glycoprotein gp120. This chapter describes the interaction of a number of glycan fragments ($(GlcNAc)_2Man_3$, Man_3, Man_5, and $GlcNAc_3$ shown schematically in

Figure 7.1-1) along with the first investigation of the DC-SIGNR CRD with a disease-associated ligand ($Man_9GlcNAc$).

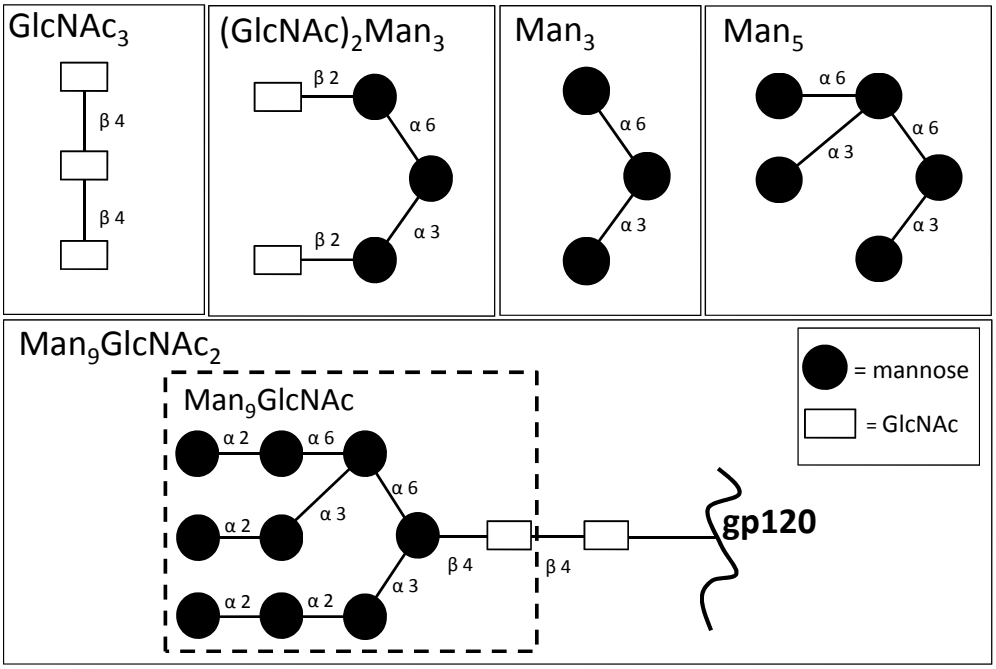

Figure 7.1-1: Glycans investigated in this study. The glycans used in this study are represented schematically. $(GlcNAc)_2Man_3$ allows comparison with the crystal structure 1K9J while $Man_9GlcNAc$ differs from the HIV associated $Man_9GlcNAc_2$ by a single, terminal GlcNAc residue.

7.1.1 Chemical Shift Perturbations in Fast and Slow Exchange.

By titrating increasing amounts of glycan into a sample of ^{15}N-labelled DC-SIGNR CRD and recording a HSQC spectrum at each titration point the chemical shift perturbations in the HSQC spectrum can be tracked. This is because the chemical shifts of a molecule in fast exchange between two states are a weighted average between those two states. In this case, exchange refers to the inter-conversion between the bound and un-bound states of the CRD. Thus, a ligand with weak affinity for the CRD will be in fast exchange. If this is the case, then increasing the concentration of ligand forces the equilibrium towards the bound state (the CRD is in the bound state more often) until eventually the chemical

shift is that of the CRD in the bound state. This is called saturation. The titration points before saturation produce chemical shift values which are a weighted average between the un-bound and bound chemical shifts resulting in a linear pattern of peaks which can be easily tracked (Figure 7.1-2).

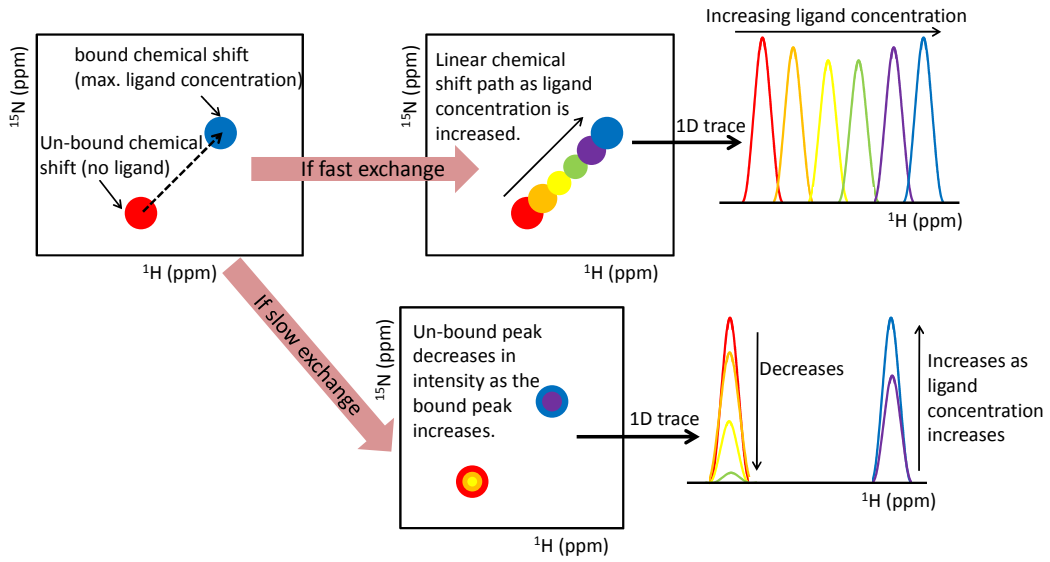

Figure 7.1-2: Illustration of chemical shift perturbations in fast and slow exchange. Ligand binding causes a change to the chemical environment of the chemical shift and so the bound chemical shift is at a different position than the un-bound. If the protein has weak affinity for the ligand then the conversion between the bound and un-bound state will be in fast exchange. In this case, as the concentration of ligand is increased chemical shifts at positions between the bound and un-bound peaks are observed. If the ligand binds more strongly, then the conversion between bound and un-bound states is slower and the protein-ligand complex enters slow exchange. As the ligand concentration is increased the un-bound chemical shift signal broadens and decreases in intensity until it disappears, then the bound chemical shift appears and increases in intensity.

The chemical shift perturbations are more complex if the CRD has a high affinity for the ligand. The protein-ligand complex enters the slow exchange regime due to the stronger binding causing slower exchange between the bound and un-bound states. In this case, the un-bound chemical shift signal broadens and decreases in intensity as the ligand concentration is increased until the signal is no longer visible. Next, the bound chemical shift signal appears and increases in

intensity with increasing ligand concentration. The chemical shifts are deemed to be saturated when the intensity of the bound chemical shift no longer increases (Figure 7.1-2). As a result, the chemical shift perturbations provide information on the lifetime of a protein-ligand complex and can be used to determine the binding affinity.

7.1.2 ^{15}N T_1 and T_2 Relaxation Parameters.

In addition to the chemical shift perturbation analyses described above measurements of transverse (T_2) and longitudinal (T_1) ^{15}N relaxation times for the holo and Man_5-bound CRD were also acquired. Relaxation is the mechanism by which the nuclear spins return to the equilibrium state. T_1 relaxation is caused by fluctuations in the magnetic field orthogonal to the z-direction, restoring thermal equilibrium, while T_2 relaxation is caused by magnetic field fluctuations in the z-direction, which results in de-phasing of the transverse magnetization. These magnetic field fluctuations are a result of the overall motion (tumbling) of the protein as well as internal motions because different orientations of the protein molecule result in changes to the electron shielding of the nuclei (due to the chemical shift anisotropy) and the strength of the dipolar coupling interactions. Therefore, the T_1 and T_2 relaxation parameters provide valuable information on the dynamics of the protein. In particular, the ^{15}N relaxation measurements provide information on the dynamics of the protein backbone (i.e. which residues are mobile and which are more rigid). These dynamics measurements are of interest due to the variable intensity of signals in the HSQC spectrum of the apo CRD (§4.2) which indicated that DC-SIGNR is very mobile. In addition, the ability

of DC-SIGNR to bind a wide range of inherently flexible glycans suggests that the dynamics of the CRD may play a vital role in glycan recognition.

7.2 Binding of Glycan Fragments with DC-SIGNR CRD.

A series of HSQC spectra were acquired upon titration of the glycan fragments $GlcNAc_3$, Man_3, Man_5, or $(GlcNAc)_2Man_3$ (shown schematically in Figure 7.1-1) into a solution of the DC-SIGNR CRD (0.7 mM) at 37˚C. All glycans induced chemical shift perturbations throughout the HSQC spectrum confirming that interaction with the CRD was taking place (HSQC spectra of fragments containing mannose are shown in Figure 7.2-1 while the HSQC spectra of the $GlcNAc_3$ titration was shown previously in §6.3, Figure 6.3-3). $(GlcNAc)_2Man_3$ (the sugar present in the 1K9J crystal structure, Figure 7.2-4) allowed the direct comparison of results in solution to published crystal data. The extent of the perturbations induced by each mannose-containing glycan ($(GlcNAc)_2Man_3$, Man_3, Man_5 and $Man_9GlcNAc$) are illustrated in Figure 7.2-2 and by $GlcNAc_3$ in Figure 7.2-3. It is clear that significant chemical shift perturbations are observed in several regions along the sequence of the CRD upon binding of all sugars. The standard deviation of the chemical shift perturbation across all residues (indicated by a dashed line in Figure 7.2-2 and Figure 7.2-3) was used (Schumann et al, 2007) to determine residues most affected by binding. Perturbations above this threshold were considered significant. For each sugar investigated, the most perturbed residues were mapped onto the DC-SIGNR CRD crystal structure obtained in complex with $(GlcNAc)_2Man_3$ (1K9J) in Figure 7.2-5.

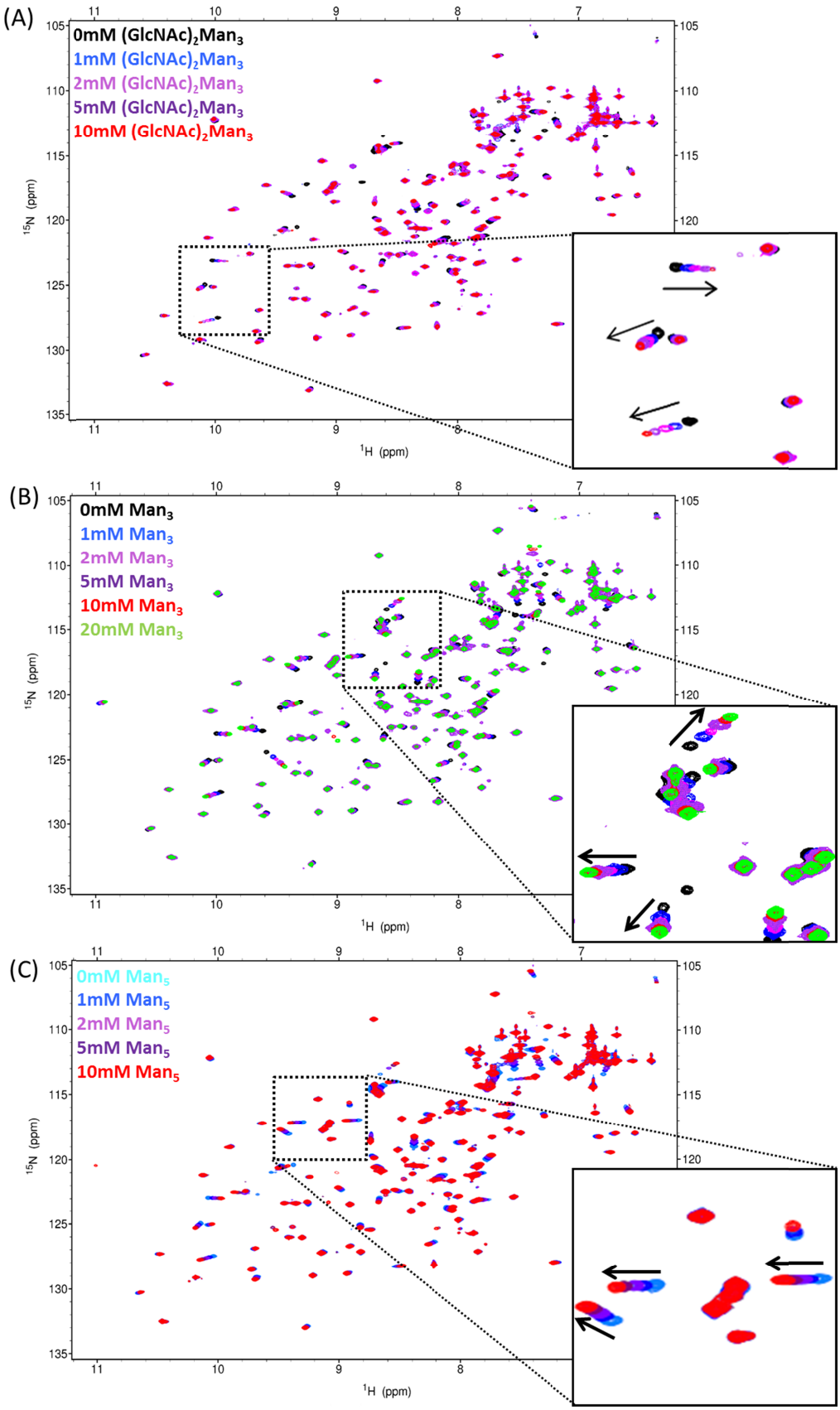

Figure 7.2-1: HSQC spectra of mannose-containing glycan fragment titrations**.** Overlaid HSQC spectra of titrations of (A) $(GlcNAc)_2Man_3$, (B) Man_3, and (C) Man_5. All three glycan fragments cause several linear chemical shift perturbations. Inset panels highlight some linear chemical shift perturbations by zooming in to the HSQC spectrum.

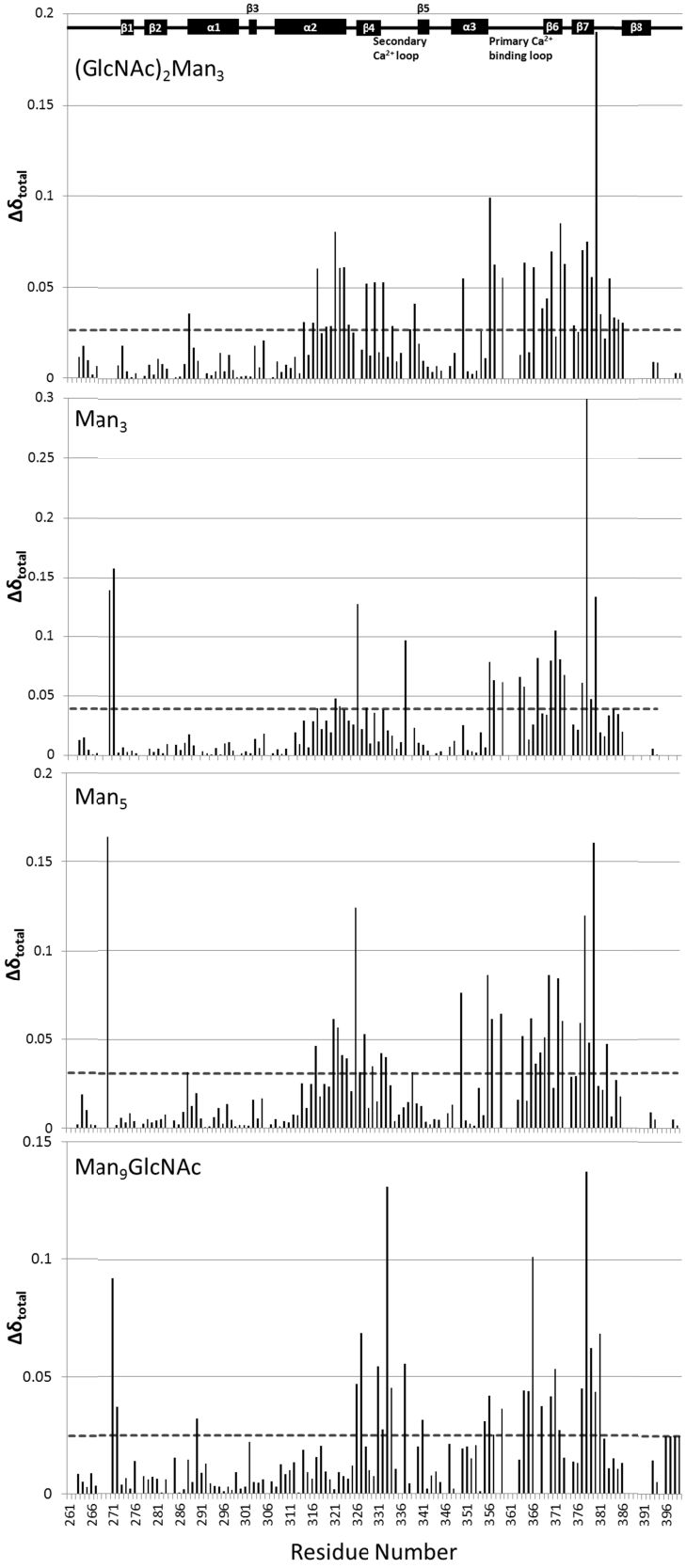

Figure 7.2-2: Chemical shift perturbations per residue for mannose-containing glycans. The total chemical shift perturbation upon addition of 5 mM of each glycan is plotted per residue. Dashed lines represent the standard deviation across all residues, perturbations above this line are considered to be significant.

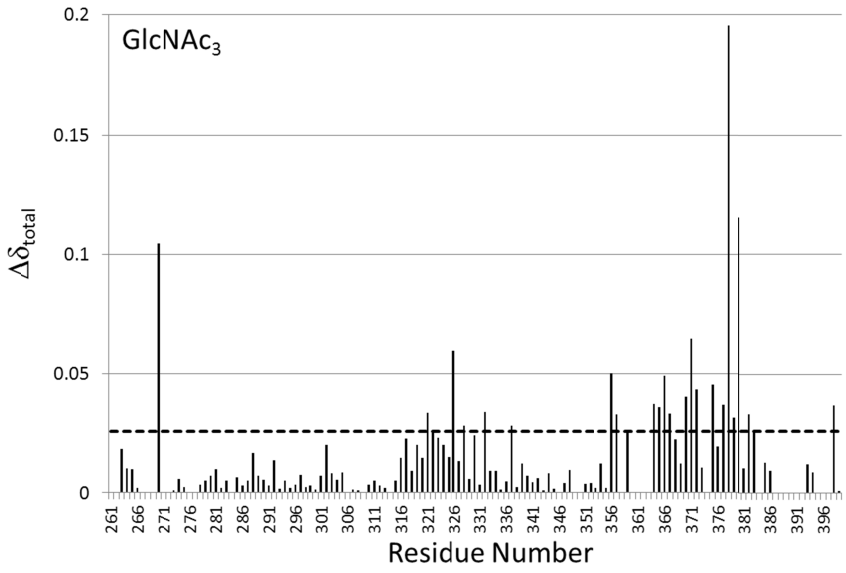

Figure 7.2-3: Total chemical shift perturbation per residue upon $GlcNAc_3$ addition. The total chemical shift perturbation upon addition of 5 mM of $GlcNAc_3$ is plotted per residue. Dashed lines represent the standard deviation across all residues, perturbations above this line are considered to be significant.

Figure 7.2-5(A-C) highlights that binding of all sugars induced perturbation of some regions in the protein near the principal glycan binding site as suggested by the 1K9J crystal structure. Specifically, residues in the loop region that make up the primary Ca^{2+}-binding site (residues 356-364) are significantly perturbed (see Figure 7.2-4 for an illustration of CRD regions of interest), supporting the suggestion that this site forms a direct interaction with the glycans (Drickamer, 1999; Feinberg et al, 2007; Feinberg et al, 2001; Guo et al, 2004). Furthermore, residues in β-sheets 6 and 7 at the core of the protein fold (residues 368-379, see Figure 7.2-4) and α-helix 2 also respond to titration of sugar.

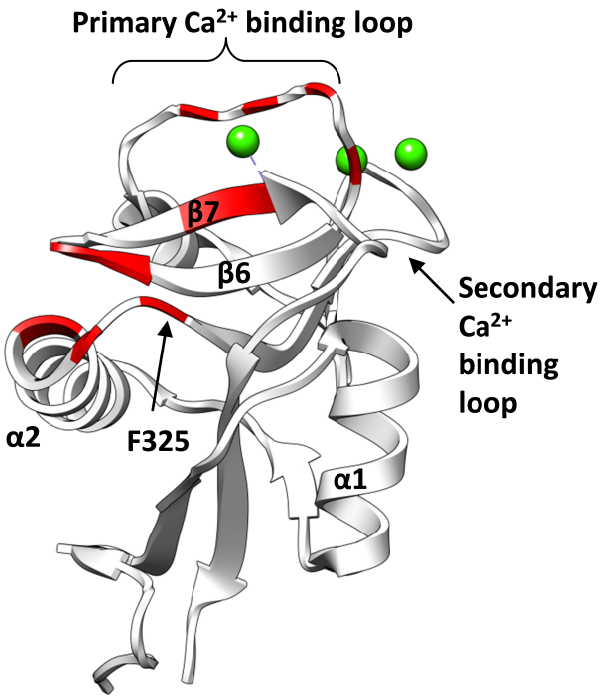

Figure 7.2-4: Residues which form contacts with $(GlcNAc)_2Man_3$ in crystal structure IK9J. Residues which form contacts with the $(GlcNAc)_2Man_3$ in the 1K9J crystal are highlighted in red and regions of interest are labelled.

7.2.1 Evidence of Different Binding Modes for the Glycan Fragments.

Despite similarities in binding to the primary calcium binding loop, the discrete glycan fragments show some marked differences in the degree of perturbations of other regions. In particular, the degree of perturbation of residues in α-helix 2 (residues 308-323), circled in Figure 7.2-5(A), vary across the glycans investigated. This is of particular interest as α-helix 2 is highlighted in the crystal structure as forming a 'shelf' complementary to the extended $(GlcNAc)_2Man_3$ sugar structure (see Figure 7.2-4 for residues implicated in binding in the 1K9J crystal structure). Only binding of $(GlcNAc)_2Man_3$, the glycan bound in the crystal structure 1K9J, yielded appreciable chemical shift perturbations in α-helix 2 (specifically residues 314-324). Even then, the F_{325} residue in α-helix 2

shown to form a direct contact with $(GlcNAc)_2Man_3$ in the crystal, was not perturbed significantly in solution. Fewer perturbations in α-helix 2 were observed upon binding of Man_3, Man_5, and $GlcNAc_3$ and no glycan induced any perturbation of the F_{325} peak. However, Man_3, Man_5, and $GlcNAc_3$ significantly perturb the residue S_{326} adjacent to F_{325} while $(GlcNAc)_2Man_3$ does not.

The chemical shift perturbations in solution also highlighted changes in regions distal to the proposed glycan binding site. Man_3 binding induced perturbation of residues 270 & 271 at the N-terminus of the CRD (solid box in Figure 7.2-5(B-C)) while $GlcNAc_3$ and Man_5 only perturbed residue 270. $(GlcNAc)_2Man_3$-binding exhibited unique perturbations in the loop region consisting of residues 382-385 (Figure 7.2-5(A), dashed box) suggesting that perturbations in this region may be due to the GlcNAc moieties. This is unusual as these residues are not involved in glycan binding seen in the crystal structure (1K9J), suggesting that the perturbations are a result of a conformational or dynamics change in this region. Interestingly, this loop region is not perturbed by $GlcNAc_3$, indicating that the size of the $(GlcNAc)_2Man_3$ molecule, in conjunction with the GlcNAc moieties, could be responsible for this unique conformational/dynamics change.

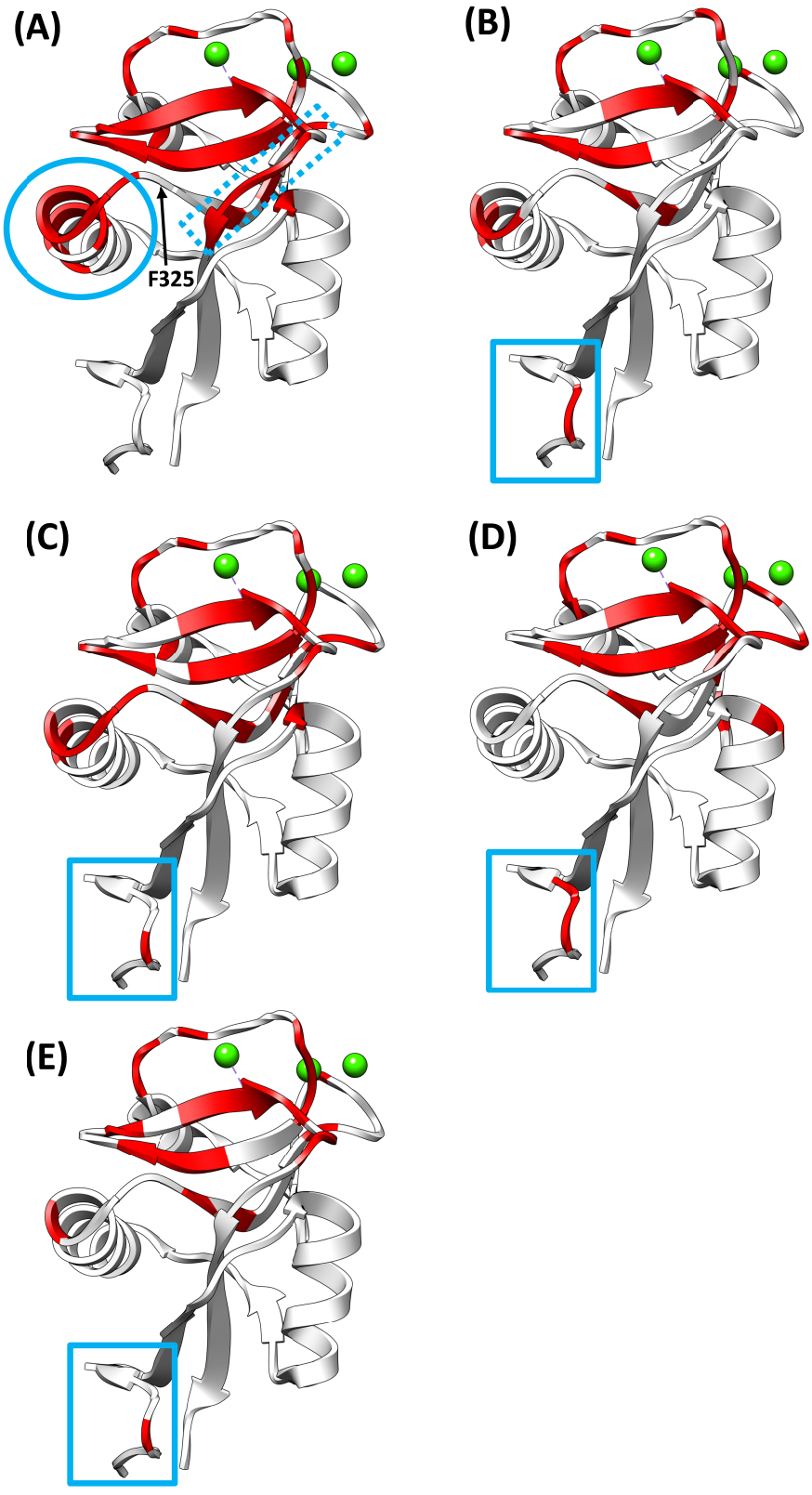

Figure 7.2-5: Significant chemical shift perturbations mapped on to 1K9J structure following addition of 5 mM (A) $(GlcNAc)_2Man_3$, (B) Man_3, (C) Man_5, (D) $Man_9GlcNAc$, and (E) $GlcNAc_3$. Residues with total chemical shift perturbations greater than the standard deviation cut-off are highlighted in red for each glycan investigated. $(GlcNAc)_2Man_3$ causes the greatest number of perturbed residues, particularly in α-helix 2 (circled) while $Man_9GlcNAc$ produces no perturbations in this helix.

7.2.2 Evidence of Glycan-Induced Conformational/Dynamics Changes.

More broadly, when these results are compared to the existing structural data, the number of chemical shifts perturbed by ligand titration in all cases is greater than was expected based on the size of the canonical binding site in the crystal and the residues implicated in these studies (see summary in Table 7.2-1). This, along with the fact that there are significant chemical shift perturbations distal to the canonical glycan binding surface shown in the crystal structure, suggests that significant changes in conformation and/or dynamics occur upon ligand binding in solution. This is in contrast to results obtained from crystallography, which yield virtually identical average structures for the free and ligand-bound states (Feinberg et al, 2001; Snyder et al, 2005a) (see Figure 7.2-6 for an overlay of two representative structures for these opposing states).

Figure 7.2-6: Overlay of glycan-bound (white) and holo (black) CRD crystal structures. All crystal structures solved of the DC-SIGNR CRD adopt nearly identical conformations. The crystal structure of the CRD in complex with $(GlcNAc)_2Man_3$ (white, 1K9J) is shown with the CRD in holo form (black, 1XPH) as an example.

Table 7.2-1: Residues significantly perturbed by ligand binding. Residues which form contacts with the $(GlcNAc)_2Man_3$ ligand in the 1K9J crystal structure are listed along with residues which are perturbed significantly upon addition of 5 mM $(GlcNAc)_2Man_3$, Man_3, Man_5, $Man_9GlcNAc$, and $GlcNAc_3$ in the NMR analyses.

crystal structure	$(GlcNAc)_2Man_3$ (5 mM)	Man_3 (5 mM)	Man_5 (5 mM)	$Man_9GlcNAc$ (5 mM)	$GlcNAc_3$ (5 mM)
	270	270	270	270	270
	271	271		271	
				272	
	288		288		
				290	
	314				
	316				
	317	317	317		
	319				
320	320				
	321	321	321		321
	322	322	322		
323	323		323		
	324		324		
325					
		326	326	326	326
			327	327	
	328	328	328		328
	330		330		
				331	
	332		332	332	332
			333	333	
	334			334	
				336	
		337		337	337
	338				
	339		339	339	
				341	
	350		350		
				355	
	356	356	356	356	356
	357	357	357	357	357
359	359	359	359	359	359
361					
363		363			
	364	364	364	364	364

				365	365
366	366		366	366	366
		367	367	367	367
	368		368	368	
	369		369	369	
	370	370	370	370	370
		371		371	371
372	372	372	372	372	372
373	373	373	373		
	375				375
377	377	377	377	377	377
378	378	378	378	378	378
	379	379	379	379	379
	380	380	380	380	380
	381			381	
	383		383		383
	384				
	385				
	386				

7.3 Binding of $Man_9GlcNAc$ to the DC-SIGNR CRD.

To extend current knowledge of the DC-SIGNR CRD to binding of more complex, physiologically important and/or disease-associated ligands, and provide a better understanding of DC-SIGNR-HIV interactions, titration experiments were carried out using $Man_9GlcNAc$ derived from the gp120 protein of HIV. $Man_9GlcNAc$ is very closely related to $Man_9GlcNAc_2$ on HIV gp120, differing by a single GlcNAc unit at the reducing terminus which would be anchored to the polypeptide backbone and hence less likely to play a crucial role in DC-SIGNR binding. As shown in Figure 7.2-2(D) (spectra are shown in Figure 7.3-1), binding of $Man_9GlcNAc$ also resulted in significant chemical shift perturbations in several regions of the CRD, and those greater than the standard

deviation are mapped onto the DC-SIGNR crystal structure (1K9J) in Figure 7.2-5(D).

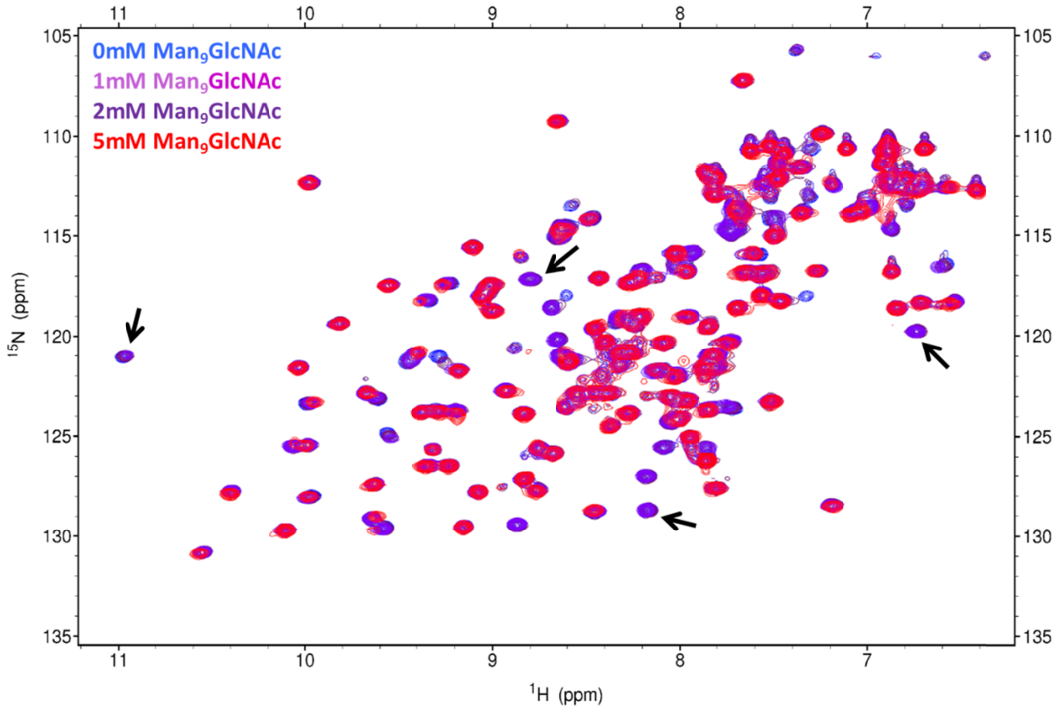

Figure 7.3-1: HSQC spectra of $Man_9GlcNAc$ titration. Titration of $Man_9GlcNAc$ into holo CRD causes some linear perturbation of chemical shifts along with line broadening. Arrows indicate some of the residues which undergo significant line broadening following $Man_9GlcNAc$ binding.

7.3.1 $Man_9GlcNAc$ has a Different Binding Mode Than the Smaller Glycan Fragments.

As observed for binding of the glycan fragments, $Man_9GlcNAc$ binding also induced perturbation of β-sheets 6 and 7 as well as the primary calcium binding loop, residues 270 and 271 at the N-terminus (solid box in Figure 7.2-5(D)) and residue S_{326} (which is also perturbed by Man_3, Man_5, and $GlcNAc_3$) below β-sheet 6. However, the perturbations induced by $Man_9GlcNAc$ show interesting differences to those induced by the smaller glycan fragments. In particular, α-

helix 2 (Figure 7.2-5), which is significantly perturbed by binding of $(GlcNAc)_2Man_3$ (and to a lesser extent by Man_5, Man_3 and $GlcNAc_3$), shows no perturbation upon binding of $Man_9GlcNAc$. In addition, residue F_{325} (Figure 7.2-4) shows no significant chemical shift perturbation upon addition of $Man_9GlcNAc$. This is interesting because the crystal structure of DC-SIGNR in complex with $(GlcNAc)_2Man_3$, along with all glycan-bound structures of the homologous protein DC-SIGN, highlight the same extended binding site, part of which is formed by α-helix 2 and F_{325} is consistently stated as an important solvent-exposed side-chain. The NMR data presented here show that, in solution, α-helix 2 is involved in binding to the small fragments (particularly $(GlcNAc)_2Man_3$) but not to $Man_9GlcNAc$. In addition, residues S_{320}, N_{323} (in α-helix 2), N_{361}, S_{363}, and G_{373} are all involved in direct sugar contacts in the $(GlcNAc)_2Man_3$ crystal structure but are not significantly perturbed on addition of $Man_9GlcNAc$. This suggests that $Man_9GlcNAc$ has a different mode of binding to the CRD than does $(GlcNAc)_2Man_3$, and highlights that inferring the mechanism of binding of larger, branched glycans from structural data of smaller, less physiologically-relevant glycans should be treated with caution. On-going work to determine a high-resolution solution structure of DC-SIGNR CRD bound to ligands will shed further light upon this.

7.4 Ligand Binding Affinities in Solution.

In addition to highlighting potential binding sites for various ligands, NMR titration data can also provide detailed affinity information. Figure 7.2-1 shows

the HSQC spectra acquired as increasing concentrations of $(GlcNAc)_2Man_3$, Man_3 or Man_5, respectively, were titrated into the DC-SIGNR CRD. The three sugar fragments behaved similarly upon titration, displaying linear chemical shift perturbations and no line broadening as ligand concentration was increased. This behaviour is characteristic of fast exchange between free and bound protein on the NMR chemical shift timescale, and suggests that the interaction between the CRD and sugar fragments is weak. This supports previous evidence that DC-SIGNR preferentially binds highly branched oligosaccharides (Feinberg et al, 2001; Guo et al, 2004; Snyder et al, 2005a). Fitting the chemical shift perturbations to a 1:1 binding model (see Materials and Methods §2.12.8) produced the best fits to the data and provided estimates of the dissociation constants for each sugar (Figure 7.4-1 shows examples of the curve fit for some residues near α-helix 2). All significantly perturbed residues in the CRD could be fitted to the same 1:1 binding model and yielded similar K_D values suggesting the presence of a single glycan binding site. Table 7.4-1 shows that $GlcNAc_3$, $(GlcNAc)_2Man_3$, Man_3 and Man_5 all bind with similar, weak affinities, yielding K_D values between 1.57-3.58 mM. This value is considerably weaker than binding of similar simple sugars to other lectin CRD domains such as galectin-1 (where K_D values for binding of lactose have been reported ranging from approximately 40 μM (Nesmelova et al, 2010) to 520 μM (Miller et al, 2009)), galectin-3 (lactose bound with a K_D = 231 μM (Diehl et al, 2010)), BclA (methyl-α-D-mannoside bound with a K_D = 2.75 μM (Lameignere et al, 2008)), and the asialoglycoprotein receptor (where binding constants of 66-539 μM were reported for a variety of simple sugars (Onizuka et al, 2012)). However, these K_D values are in line with the affinity of the analogous protein DC-

SIGN for other small glycans such as fucose (6 mM) and mannose (13 mM) (Timpano et al, 2008).

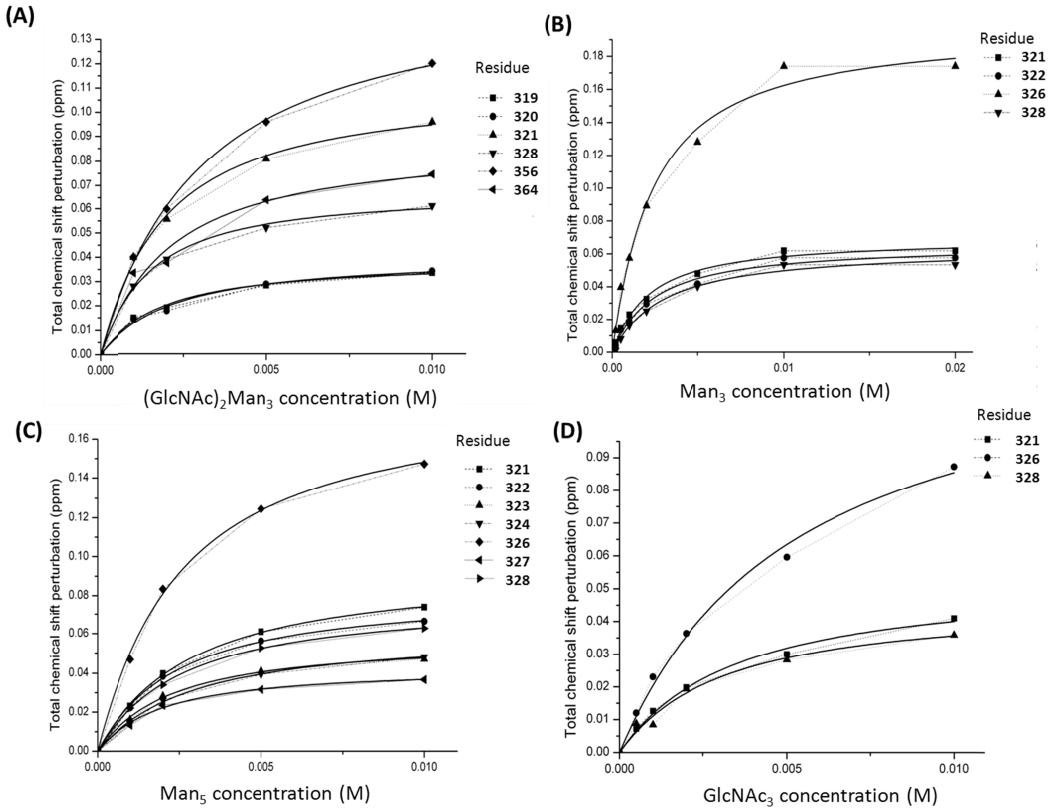

Figure 7.4-1: Chemical shift perturbations fit to a single-site binding model for (A) $(GlcNAc)_2Man_3$, (B) Man_3, (C) Man_5, and (D) $GlcNAc_3$. The total chemical shift perturbation per residue is plotted against glycan concentration and fit to a single-site binding model. All significantly perturbed residues fit the same single-site model but here only residues in α-helix two are shown as an illustration. The fit was used to calculate the dissociation constant of the CRD for each glycan fragment.

Table 7.4-1: Dissociation constants calculated from chemical shift perturbations. Disassociation constants (K_D) were calculated for each glycan fragment by fitting all residues significantly perturbed to a 1-1 binding model. The K_D values show that DC-SIGNR binds the glycan fragments weakly and has similar affinity for each sugar.

Glycan	K_D (mM)
$(GlcNAc)_2Man_3$	1.57 ± 0.46
Man_3	2.04 ± 0.54
Man_5	2.20 ± 0.43
$GlcNAc_3$	3.58 ± 0.66

7.4.1 The DC-SIGNR CRD has Higher Affinity for $Man_9GlcNAc$ than the Glycan Fragments.

In contrast to the glycan fragments described above, we can see from Figure 7.3-1 that binding of $Man_9GlcNAc$ causes significant broadening and disappearance of a number of CRD resonances. This is consistent with intermediate exchange on the NMR time scale (Keeler, 2005; Lian & Roberts, 2011). This implies that $Man_9GlcNAc$ binds with higher affinity than $(GlcNAc)_2Man_3$, Man_3, $GlcNAc_3$ or Man_5, since we move from fast exchange to intermediate exchange on the NMR chemical shift timescale as the lifetime of a complex is increased. This was expected and is consistent with previous competition assay data obtained for DC-SIGNR binding to $Man_9GlcNAc_2$ which yielded a K_i value of 200 μM (Mitchell 2001), as well as other lectins (such as galectin-1) that show increased affinity for larger, more complex glycans (Miller et al, 2009). Due to the severe line broadening at the highest $Man_9GlcNAc$ concentrations (5 mM, a concentration estimated to saturate binding at 25× the reported K_i) resulting from intermediate-timescale exchange, an accurate dissociation constant could not be estimated for $Man_9GlcNAc$ binding. However, it is generally accepted that intermediate exchange corresponds to dissociation constants in the range of 10^{-3}-10^{-7} M (Lian & Roberts, 2011).

7.5 NMR Dynamics Reveal a High Degree of Flexibility for the Ca^{2+}-Bound (Holo) DC-SIGNR-CRD.

Because a larger number of chemical shift perturbations were observed than expected based on the size of the canonical binding site present in the CRD crystal structure, residue by residue changes in ^{15}N relaxation times upon glycan binding were also investigated. The high degree of dynamics in the DCSIGNR CRD was first suspected because of the variable signal intensities in the HSQC spectrum of the apo CRD (§4.2). These variable signal intensities suggest that, in its Ca^{2+}-free form, the protein can exchange between an ensemble of conformational states with a rate corresponding to the intermediate exchange regime on the NMR ^{1}H and ^{15}N chemical shift timescale at 700 and 800 MHz.

To explore the dynamics of the CRD further, T_1 and T_2 relaxation times were measured for the CRD in the absence (holo) and presence of 10 mM Man_5 (Figure 7.5-1 (A) and (B), and Table 7.5-1). Man_5 was selected for this study because binding of $Man_9GlcNAc$ produced severely exchange-broadened spectra, preventing accurate measurement of relaxation parameters. The T_1 data are very similar for the two states, showing a similar trend across the protein, with average values of 788.26 $\pm$ 22.4 ms for the holo and 806.22 $\pm$ 40.35 ms for the Man_5-bound forms. However, more significant differences are observed in the transverse relaxation time constants (T_2). For the holo CRD, though most of the T_2 values fall near the average (76.27 $\pm$ 2.27 ms), residues in both Ca^{2+} binding loops, α-helix 1, β-sheets 6 and 7, and at the N-terminus (see shaded regions in Figure 7.5-1) display shorter relaxation times suggesting either more structured

conformations in these regions or chemical exchange processes (or both) (Csizmok et al, 2008). A similar trend is seen in the T_2 data for the Man_5-bound CRD, but with an overall lower average T_2 relaxation time (66.54 $\pm$ 8.23 ms) and further reductions in the T_2 transverse relaxation times observed in all regions listed above.

7.5.1 *Man_5 Binding Hinders the Motion of the Primary Calcium Loop.*

Asterisks in Figure 7.5-1 indicate residues in the Ca^{2+} binding loops whose signals are so severely broadened as to be unobservable in these experiments, supporting the rapid relaxation of these regions. Specifically, these residues included E_{359}, N_{361} and N_{362} in the primary Ca^{2+} binding loop, which make up the EPN motif conserved among all mannose-binding C-type lectins (Drickamer, 1992). Binding at the primary Ca^{2+} site is well characterised (it is a distinguishing feature of C-type lectin binding) and it has been confirmed that the EPN sequence is responsible for mannose specificity (Drickamer, 1992). In addition to the EPN motif, residues across the entire primary Ca^{2+} binding loop display enhanced transverse relaxation in both the holo and ligand bound forms. For the holo form, motions on the μs-ms timescale near the primary Ca^{2+} site may be due to conformational exchange as the Ca^{2+} ions come on and off the CRD. There is previous evidence for several other C-type lectins that, in the absence of bound calcium, undergo *cis-trans* isomerisation about the peptide bond of the conserved proline (P_{360}) in the EPN motif (Ho et al, 2010; Ng et al, 1998; Nielbo et al, 2004; Pavlicek et al, 2003; Poget et al, 2001). In most cases, binding of Ca^{2+} has resulted in stabilization of one conformer; however no data for DC-SIGNR

have been reported thus far in this regard. The data described in this study suggest that considerable local dynamics persist in the Ca^{2+} bound form. The further reduction in T_2 values upon addition of Man_5 suggests a further reduction in the rate of dynamic motion (i.e. motion is more hindered) in the primary calcium binding loop upon sugar binding, thus supporting previous reports that that this is the site of key CRD-mannose interactions.

Table 7.5-1: Average relaxation parameters of holo and Man_5-bound CRD. Average T_1, T_2 and the ratio of T_1/T_2 were calculated across all residues. The rotational correlation time (τ_c) was estimated from the T_1/T_2 ratio excluding residues with ratios greater/less than one standard deviation above/below the mean.

	Holo CRD	**Man_5 bound CRD**
Average T_1 (ms)	788.26 ± 22.4	806.22 ± 40.35
Average T_2 (ms)	76.27 ± 2.27	66.54 ± 8.23
Average T_1/T_2	10.67 ± 0.43	13.19 ± 1.75
Rotational correlation time (τ_c) (ns)	10.4 ± 0.4	12.5 ± 1.36

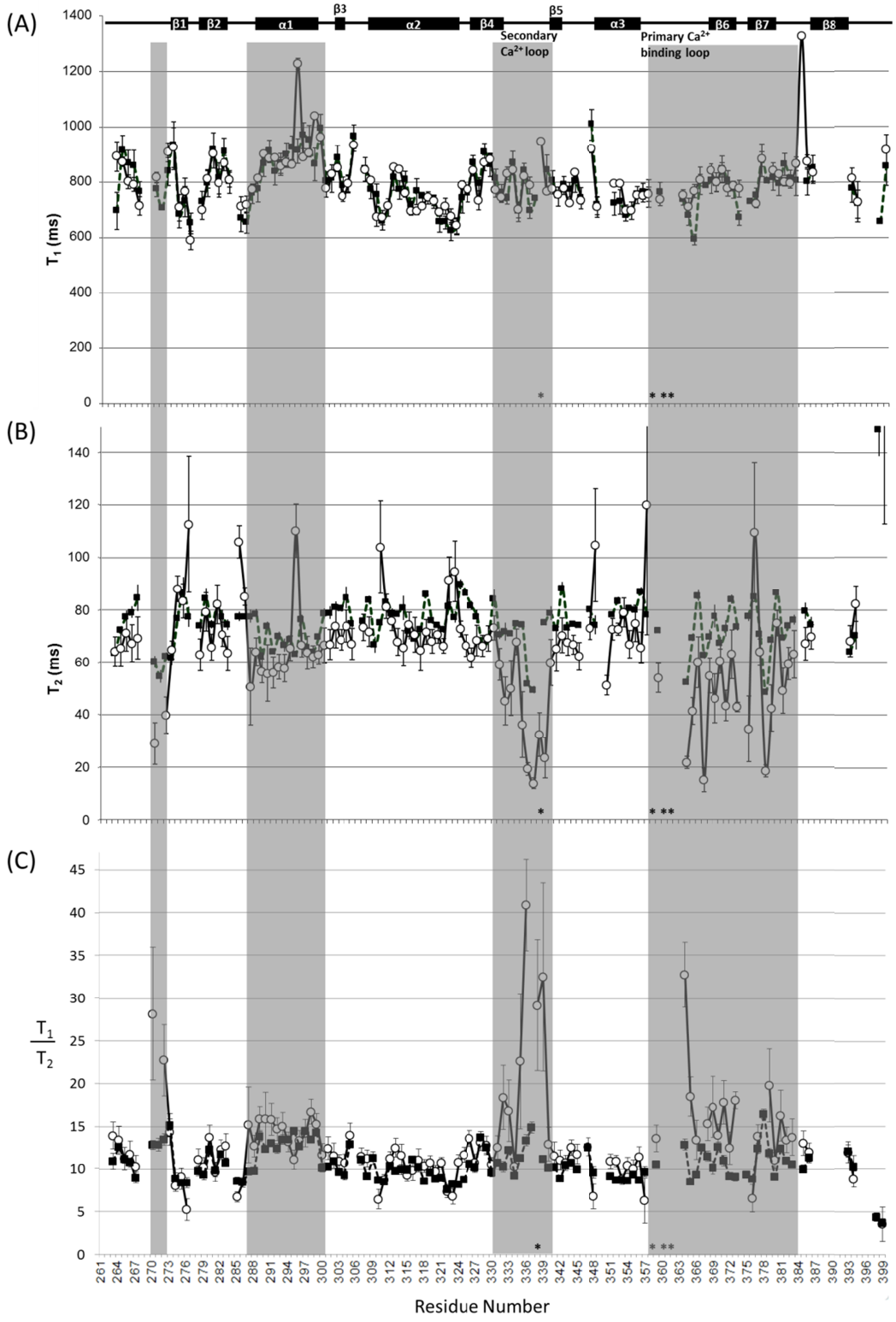

Figure 7.5-1: ^{15}N T_1 and T_2 dynamics of holo (solid square) and Man_5-bound (circle) DC-SIGNR CRD. Plot of per residue values of (A) ^{15}N T_1, (B) ^{15}N T_2, and (C) ^{15}N T_1/T_2 for holo (solid squares) and Man_5-bound (open circles) DC-SIGNR CRD. Asterisks denote residues which are not observed due to fast relaxation or exchange.

A decrease in T_2 upon addition of Man_5 is also observed for residues throughout β-sheets 6 and 7 which, as discussed earlier, form part of the sugar binding site proposed in the literature. Specifically, it is thought that the loop joining these two strands directly interacts with the glycan, thus sugar binding to the loop region may hinder the motions of the individual strands. This hindered motion would lead to enhanced transverse relaxation as the rate of the dynamics (both global and local) changes in this region.

7.5.2 Dynamics Changes Distal to the Ligand Binding Site.

The relaxation data presented here also highlighted regions in the CRD that have not yet been implicated in binding to sugars, namely the secondary calcium binding loop and α-helix 1. Residues all along the length of α-helix 1 show a subtle but significant reduction in T_2 in the holo CRD, and a further reduction upon binding of Man_5. α-helix 1 is positioned towards the N-terminus of the CRD, where increased transverse relaxation rates are also observed for residues 269-272. The N-terminus of the CRD connects to the α-helical neck in the full length protein, and others have proposed that this region forms a flexible 'hinge' (Leckband et al, 2011; Snyder et al, 2005a; Yu et al, 2009) allowing the CRD to sample multiple orientations with respect to the neck. Taking this previous work into account, one tentative explanation for the enhanced relaxation in α-helix 1 and the N-terminus is that binding of Man_5 restricts/hinders motions at the flexible 'hinge', locking the CRD into a more restricted set of interconverting conformations with respect to the neck domain. Given the size of the Man_5 ligand, the linear chemical shift perturbations

obtained, and the fit to a 1:1 binding model, it is unlikely that Man_5 is binding to multiple sites in the CRD, but it is likely that the binding of the glycan has a marked effect on the dynamics of regions distal to the binding site.

The secondary calcium binding site also responds to glycan binding, despite the fact that this region has not previously been shown to play a role in direct glycan contact within any crystal structure. In the holo protein, the reduced T_2 values in this region (T_{337} was broadened beyond detection) were attributed to slower internal motions of the loop (on the millisecond to microsecond timescale) due to binding of Ca^{2+}. This fits with the interpretation of the relaxation data for the primary Ca^{2+}-binding loop described above. The further enhancement in transverse relaxation rates upon Man_5 binding is less obvious, since this region of the CRD has not been previously shown to form part of the extended glycan binding site. The secondary calcium loop lies in close proximity to the proposed binding site, and given the structural plasticity proposed for C-type lectins which allows them to accommodate a wide range of diverse ligands, it is possible that Man_5 has a different mode of binding to the DCSIGNR CRD than was shown for $(GlcNAc)_2Man_3$ which includes the secondary Ca^{2+} binding loop.

7.5.3 Rotational Correlation Time (τ_c)

The average T_2 value decreased from 76.27 to 66.54 ms upon ligand binding. This is attributed to slower tumbling of the ligand-bound protein compared to the holo protein in solution. This was confirmed by using the T_1/T_2 ratio (Figure 7.5-1(C)) as an estimate of the overall rotational correlation time (τ_c)

by first excluding residues with T_1/T_2 values more than one standard deviation from the average (and thus experience a significant contribution from either chemical exchange or internal motion (Clore et al, 1990)). A value for the relaxation-derived τ_c of 10.4 $\pm$ 0.4 ns was obtained for the holo CRD, which is only slightly longer than the expected value of 8.55 ns obtained using the general rule of 0.5 ns τ_c per 1 kDa molecular weight (Clore et al, 1990; Copie et al, 1996; Kay et al, 1989). This deviation from the ideal value is not large enough to infer oligomerization of the CRD, but may reflect a non-spherical shape of the monomeric protein. There is a modest (~17%) increase in the rotational correlation time from 10.4 to 12.5 ns upon binding of Man_5 to the CRD (Table 7.5-1). The addition of a single Man_5 molecule is unlikely to cause any appreciable increase in the rotational correlation time, and binding induced aggregation would result in a much larger increase in τ_c, suggesting that the CRD adopts a more 'open' conformation as a result of Man_5 binding. This could be due to the carbohydrate binding site 'opening' to accommodate Man_5 (however, this is not seen in the crystal data for other sugars). A more 'open' structure for the ligand bound state has also been observed for the CRD of galectin-1 (Nesmelova et al, 2010).

In broad terms, comparison of the relaxation and chemical shift perturbation data demonstrate that while several regions in the CRD undergo marked changes in dynamics upon glycan binding, many more residues display chemical shift perturbations. As a result all of the chemical shift perturbations throughout the CRD cannot be explained by dynamics changes alone. Therefore ligand-induced conformational changes must also play a role. The extent of these

conformational changes warrants further investigation, as no conformational changes have been observed in any published crystal structures for DC-SIGNR CRD upon ligand binding. Future work will involve solving the solution structures of holo- and ligand-bound states in an effort to characterise these conformational changes.

7.6 Conclusions.

This chapter describes the first solution state NMR studies of glycan binding to the DC-SIGNR CRD along with the first structural information on binding to a large, disease-associated ligand: $Man_9GlcNAc$. Chemical shift perturbation studies allowed investigation of a range of glycans in solution and provided several interesting observations. The glycan fragments ($GlcNAc_3$, $(GlcNAc)_2Man_3$, Man_3, and Man_5) investigated all displayed significant chemical shifts in the primary calcium binding loop, along with residues in β-sheets 6 and 7, supporting previous research which shows that this region (and particularly the EPN motif in the primary calcium binding loop) is responsible for mannose-specific binding of C-type lectins. However, the perturbation data also shows that the three glycan fragments have different, extended binding modes. Most interestingly, α-helix 2 was perturbed most by $(GlcNAc)_2Man_3$ (supporting the 1K9J crystal structure which shows that α-helix 2 forms a 'shelf' complementary to the glycan structure) but to a lesser extent by $GlcNAc_3$, Man_3 and Man_5. In addition, the loop region formed by residues 382-385 was only perturbed by $(GlcNAc)_2Man_3$ suggesting that this glycan may have a significantly different

binding mode than Man_3 and Man_5, possibly due to the GlcNAc moieties. However, the lack of perturbation in this region by $GlcNAc_3$ suggests that the size of the $(GlcNAc)_2Man_3$ and, possibly, the position of the GlcNAc moieties may also play a role in the conformational change of this loop region. The unique perturbations in this loop region are likely due to conformational/dynamics changes because there are no direct contacts with $(GlcNAc)_2Man_3$ in the crystal structure. In addition, $(GlcNAc)_2Man_3$ causes chemical shift perturbations throughout the CRD, many more than was expected based on the size of the ligand and the canonical binding site in the crystal. This could be because $(GlcNAc)_2Man_3$ is not a 'natural' ligand of DC-SIGNR (it has not been highlighted in any glycan-array screenings suggesting that DC-SIGNR has very weak affinity for $(GlcNAc)_2Man_3$ (Guo et al, 2004; Yabe et al, 2010)) therefore the CRD must undergo more extensive conformation/dynamics changes in order to accommodate it. This is in contrast to $Man_9GlcNAc$ binding, which caused fewer chemical shift perturbations (and in different regions of the CRD) despite it being larger.

7.6.1 $Man_9GlcNAc$ has a Different Binding Mode than the Glycan Fragments.

$Man_9GlcNAc$ displayed significantly different chemical shift perturbations than the glycan fragments investigated. Most notably, no significant chemical shift perturbations occur in α-helix 2 despite this helix forming the extended binding site in the 1K9J crystal structure, and all glycan-bound structures (of small glycan fragments) of the analogous protein DC-SIGN. This suggests that

$Man_9GlcNAc$ has a significantly different binding mode than the glycan fragments and that the use of small glycans as models for binding of larger, physiological ligands should be treated with caution. In addition, this study of the DC-SIGNR CRD-$Man_9GlcNAc$ complex demonstrates that solution state NMR is potentially more accommodating, informative and essential for the design of drug molecules that could inhibit binding of large, disease associated carbohydrates and glycoproteins.

The NMR data presented here also allowed the first estimates of the binding affinities of the glycan fragments for the DCSIGNR CRD. All three glycan fragments bound with similar, weak affinities (K_D = 1.57-3.5 mM). The affinity of $Man_9GlcNAc$ for the CRD was significantly higher, as indicated by severe signal broadening unique to this ligand and characteristic of intermediate exchange on the NMR timescale (which corresponds to K_D values in the range 10^{-3}-10^{-7} M). This higher affinity could explain why $Man_9GlcNAc$ (the largest of the ligands tested) yielded the smallest number of chemical shift perturbations: the binding site in the DCSIGNR CRD is "tuned" for this ligand and does not need to rearrange significantly in order to accommodate it.

7.6.2 The DC-SIGNR CRD is Very Dynamic Even Following Ca^{2+} and Man_5 Binding.

The chemical shift perturbation data was extended by the first investigation of the DC-SIGNR CRD dynamics in both the holo and Man_5-bound form. The amount of variable intensity in the HSQC spectrum was reduced upon addition of calcium. Residues in the primary calcium binding loop were

particularly affected by calcium addition suggesting that binding of calcium alters the conformational dynamics of the protein. It is possible that the CRD converts between multiple conformations in the apo form (causing the line broadening seen in the apo HSQC spectrum) and that calcium binding hinders the conversion between these conformations. This is supported by the dynamics data which shows that the holo CRD displays hindered dynamics in the calcium-binding loops as well as α-helix 1 and residues at the N-terminus. This indicates that (a) there may be some stabilization of structure in these regions as a result of calcium binding, (b) the holo CRD is sampling fewer conformations than the apo CRD, and/or (c) the rate of interconversion between the conformations is slower (more hindered). However, despite the stabilizing effect of calcium binding, the holo CRD maintains a level of mobility.

Interestingly, binding of Man_5 to the DC-SIGNR CRD did not completely diminish the dynamics, and interconversion (albeit on a slower timescale) still prevailed. However, the average T_2 value was reduced upon Man_5 binding denoting slower overall global rotational diffusion (larger τ_c) and reflecting slightly slower internal motions. This would suggest that sugar-binding further hinders the local internal motions (and the global tumbling), thus further reducing the overall flexibility of the protein. This is most pronounced in both calcium-binding loops, β-sheets 6 and 7, α-helix 1 and the N-terminus. These results support previous reports which show that the primary calcium binding loop and β-sheets 6 and 7 are sites of key CRD-mannose interactions, while

highlighting the secondary calcium-binding loop, α-helix 1 and the N-terminus as potentially important new sites.

Binding of $Man_9GlcNAc$ was again unique in that these spectra displayed severe line-broadening, which could be interpreted as further hindered dynamics in the CRD (compared with Man_5–associated CRD). It follows that the larger, higher-affinity $Man_9GlcNAc$ could restrict/slow the motions of the CRD more than the smaller, low affinity Man_5, and that the CRD would interconvert between a smaller number of possible conformations when this ligand was bound.

7.6.3 Ligand Binding Induces Dynamics/Conformational Changes Distal to the Glycan Binding Site.

Analysis of the backbone relaxation (and chemical shift perturbation) data showed that Man_5 binding not only had an effect on the dynamics of regions in the CRD near the proposed glycan binding site (primary calcium-binding loop, β-sheets 6 and 7), but also had a pronounced effect on distal regions of the CRD. On addition of Man_5, the N-terminal half of α-helix 1 along with residues at the N-terminus (most pronounced for D_{271}) show evidence of hindered dynamics. This region of the CRD forms the link between the CRD and the helical neck region in full length DC-SIGNR, and decreased mobility in this region suggests that ligand induced signals could be propagated through the CRD, along the neck, to the cytoplasmic region of the full-length protein. Alternatively, ligand binding could alter the orientation of the CRD with respect to the neck promoting multi-valent binding by adjacent CRDs.

It is clear from the chemical shift perturbation data that more residues are affected by ligand binding than can be explained by direct interaction of the protein with the oligosaccharides. This implies that the DCSIGNR CRD is a highly flexible, dynamic domain that can interconvert between multiple conformations. Further work is needed to confirm the extent of these ligand-induced effects, for which solution NMR is ideally suited. On-going determination of the solution structure of both the holo- and ligand-bound CRD along with saturation transfer difference experiments and further backbone relaxation measurements could further characterise the conformational and dynamics effects of glycan binding. In addition, the acquisition of heteronuclear NOE experiments would provide a more detailed analysis of the protein dynamics using the model-free formalism described by Lipari-Szabo (Lipari & Szabo, 1982a; Lipari & Szabo, 1982b). In the future, this work can easily and quickly be extended to investigate more ligands including disease-associated glycoproteins and potential drug molecules.

8 Conclusions & Future Work.

Complex carbohydrate binding events that occur within the human immune system are absolutely vital to healthy immune function and to proper host responses to a wide variety of pathogens and parasites. Greater understanding of this essential glycoimmunology field promises to provide important insights into major world health risks such as HIV, tuberculosis and Gram-negative multiresistance - diseases that remain incurable and possess salient structural carbohydrates that are central to their infectious cycles. C-type lectins in serum and on leucocytes and endothelial cells represent some of the most important receptors for complex carbohydrates and their roles in contributing to sophisticated pathogen recognition and cellular response mechanisms are only just beginning to emerge. Structural studies have provided insights into the mechanisms via which C-type lectins assemble and bind to their targets with specificity, displaying a range of strategies including oligomerization, monosaccharide selectivity and, in some cases, extended binding sites incorporating multiple protein-glycan contacts across broader surfaces of both lectin domain and oligosaccharide. However, characterisation of the interaction of larger, disease-associated glycans with C-type lectins (such as HIV-derived

$Man_9GlcNAc$ with the human C-type lectin DC-SIGNR) has not been reported thus far.

The C-type lectin DC-SIGNR is particularly worthy of study due to its ability to bind a wide range of disease-associated oligosaccharides via its carbohydrate recognition domain (CRD) and promote both *incis* and *intrans* infection. Several crystal structures have been solved of the DC-SIGNR CRD which provide valuable insight in to the protein conformation and extended binding sites responsible for interactions with small glycan fragments such as $(GlcNAc)_2Man_3$ (namely the 'shelf' region formed by α-helix 2 and the solvent exposed F_{325} residue). However, structures of the DC-SIGNR CRD in complex with larger, disease-associated glycans remain elusive. As a result, the interactions of the DC-SIGNR CRD with oligosaccharides such as $Man_9GlcNAc_2$, present on the HIV envelope glycoprotein gp120, are assumed to be analogous to that of the $(GlcNAc)_2Man_3$ glycan present in the 1K9J crystal structure. In addition, the calcium binding properties and dynamic nature of the CRD have not been investigated.

Here the first solution state NMR study of the DC-SIGNR CRD has been described. A protocol for the production of uniformly double labelled ($^{15}N/^{13}C$) DC-SIGNR CRD in abundance has been successfully developed and the backbone and side-chain assignment of the CRD in the holo form has been reported. The assigned HSQC has provided a 'map' of the protein backbone and allowed the characterisation of the binding properties of calcium ions and a variety of glycan ligands (including $Man_9GlcNAc$) with the DC-SIGNR CRD in solution. The pH dependence of calcium and glycan binding has also been investigated. In addition,

solution NMR methods have been used to initiate characterisation of the molecular dynamics of the CRD in its free and ligand-bound states. These data have revealed several interesting properties of the DC-SIGNR CRD which are summarized below.

8.1.1 Calcium Binding Induces Significant Conformational/Dynamics Changes in the DC-SIGNR CRD.

The data presented here suggest that the DC-SIGNR CRD is highly dynamic and may convert between multiple conformations in the apo state. The addition of calcium causes a large number of chemical shift perturbations throughout the HSQC spectrum suggesting a global conformation/dynamics change in the CRD. These changes are most pronounced in the primary calcium binding loop, where signals for residues N_{361} and N_{362} are absent in the apo HSQC spectrum. The calcium induced conformational and dynamics changes could be responsible for the glycan binding ability of the CRD, possibly by formation of a glycan binding site or by hindering the inter-conversion between conformational states so that conformations capable of binding ligands are favoured. However, further work is needed in order to characterise these calcium induced changes.

EDTA titration experiments revealed that the DC-SIGNR CRD has very high affinity for calcium. The fact that a calcium-depleted HSQC spectrum is acquired in the presence of excess EDTA (as opposed to a spectrum identical to the apo HSQC spectrum) suggests that the calcium binding sites in the CRD may have different affinities. This supports previous work which has shown that other C-type lectins, such as the asialoglycoprotein receptor, possess both high and low

affinity calcium binding sites (Onizuka et al, 2012). The use of techniques such as isothermal titration calorimetry could help in determining the CRD calcium affinity in the future.

8.1.2 The DC-SIGNR CRD Releases Ligand at Low pH.

Comparison of HSQC spectra of both holo and glycan-bound DC-SIGNR CRD at pH 6.8 and 4.2 suggest that the DC-SIGNR CRD releases ligand (and calcium) at low pH. This supports previous suggestions that the DC-SIGNR CRD can act as an endocytic receptor in certain cells, internalising ligands and releasing them in a low pH endosome environment. While the HSQC spectra of holo and glycan bound CRD at low pH are extremely similar, there are subtle differences suggesting glycan binding is not completely abolished at low pH values. Eluting protein bound to a mannose-Sepharose column using pH 4.0 buffer confirmed that the CRD releases glycan at low pH and that the glycan binding affinity is not completely abolished (as protein eluted across more fractions than when using the EDTA elution method). Future work is required to confirm the extent of this pH dependence. In particular, investigation of the apo CRD at low pH and the effect of pH on mannose-containing ligands (particularly disease-associated glycans) is required.

8.1.3 Different Binding Modes for Small Glycans Versus $Man_9GlcNAc$.

Interestingly, members of the C-type lectin family contain nonregular secondary structure in the CRD and, typically, the ligand binding sites are located in these nonregular regions. Another key feature of the C-type lectin family is its enormous potential for ligand binding diversity, despite the majority of C-type

lectin proteins containing the same CRD scaffold. This ligand binding diversity is due to variety in the non-regular loop regions across C-type lectin family members (Drickamer & Taylor, 2005). It is very likely that the C-type lectin family has evolved into a range of homologous proteins (with the same scaffold) with a very broad spectrum of ligand specificities, including targets of both exogenous and endogenous origin.

The different binding modes for the glycans studied here, as indicated by the unique patterns of chemical shift perturbations, support the structural plasticity proposed for C-type lectins, which allows them to accommodate a wide range of diverse ligands. All glycans investigated in this study caused perturbation of protein regions near the principal glycan binding site as suggested by the 1K9J crystal structure, namely the primary Ca^{2+}-binding loop and β-sheets 6 and 7. However, each glycan had a unique set of additional perturbations in α-helices 1 and 2, the N-terminus of the CRD, and the loop region consisting of residues 382-385. For example, the majority of residues in α-helix 2 were preferentially engaged during binding of $(GlcNAc)_2Man_3$, which supported its role as a critical region (along with residue F_{325}) in forming a 'shelf' complementary to the $(GlcNAc)_2Man_3$ structure in the extended glycan binding site. However, fewer residues were affected in α-helix 2 upon binding of $GlcNAc_3$, Man_3 and Man_5, and none were affected by $Man_9GlcNAc$ binding. This suggests that, while the primary binding mode (including the primary Ca^{2+} loop and β-sheets 6 and 7) is similar for each glycan, the extended binding sites/modes differ. In addition, the $(GlcNAc)_2Man_3$ glycan causes the largest number of chemical shift perturbations overall, far more than expected based on the size of the canonical binding site

present in the 1K9J crystal structure. This could be a result of $(GlcNAc)_2Man_3$ not being a 'natural' ligand of DC-SIGNR (as compared to branched oligomannose structures) therefore the CRD must undergo more extensive conformation/dynamics changes in order to accommodate it. This is in contrast to $Man_9GlcNAc$ binding, which caused fewer chemical shift perturbations (and in different regions of the CRD) despite it being larger.

The NMR data presented here also provided the first quantitative estimates of the direct binding affinities of the small glycan fragments investigated for the DCSIGNR CRD. All three glycan fragments bound with similar, weak (K_D = 1.57-2.2 mM) affinities. The affinity of $Man_9GlcNAc$ for the CRD was significantly higher, as indicated by severe signal broadening unique to this ligand and characteristic of intermediate exchange on the NMR timescale (which corresponds to K_D values in the range 10^{-3}-10^{-7} M). This higher affinity helps explain the fact that $Man_9GlcNAc$ (the largest of the ligands tested) yielded the smallest number of chemical shift perturbations, i.e. the binding site in the DCSIGNR CRD is "tuned" for this ligand and does not need to rearrange significantly in order to accommodate it.

Overall, the data suggest that use of small glycans as models for illustrating the binding of larger, physiological ligands by lectins such as DC-SIGNR should be treated with caution. This strongly supports the notion that direct and empirical investigations of the interactions of large, disease-associated ligands with the DC-SIGNR CRD are required to provide future understanding of the basic function of these molecules and also drive the identification and development of

effective drug molecules that could inhibit these interactions. NMR is ideally suited for this task as it allows the investigation of large, disease-associated ligands in solution.

8.1.4 The DC-SIGNR CRD is Very Dynamic.

Whilst several structural analyses of mammalian C-type lectins have revealed substantial spatial information on glycan ligand binding, the level of dynamics data relating to these carbohydrate-binding proteins is surprisingly limited. The fact that members of the C-type lectin family typically possess the same CRD scaffold, and yet display a wide range of ligand binding specificities and affinities could mean that the dynamics of these molecules, especially in their nonregular regions, play important roles. Previous studies on the tunicate C-type lectin TC14 have shown that the nonregular sequences in the CTLD are rigid (Poget et al, 2001). However, just as primary sequences and ligand specificity for the C-type lectin family are many and varied, so too could the dynamic properties of the assorted domain family members. Indeed, other C-type lectins have been reported to exhibit flexibility in the nonregular calcium binding loops (Sallum et al, 2007; Wurzburg et al, 2006).

The data presented here indicate that, contrary to C-type lectins such as TC14, DC-SIGNR shows considerable flexibility within its nonregular sequences and this may contribute to its ability to interact with large, flexible glycans and transduce intracellular signals. In its Ca^{2+}-free form, the apo CRD appeared to be very dynamic, probably exchanging between an ensemble of conformations. The holo DC-SIGNR CRD displays hindered dynamics in the calcium-binding loops as

well as α-helix 1 and residues at the N-terminus, indicating that there may be some stabilization of structure in these regions, the holo CRD is sampling fewer conformations than the apo CRD, and/or the rate of interconversion between the conformations is slower (more hindered). However, the CRD maintains a degree of mobility in the presence of calcium suggesting that the holo CRD still samples multiple conformations. An explanation for this mobility may be the exchange of Ca^{2+} ions on and off the protein (but more work is required to confirm this).

Interestingly, the dynamics are not diminished completely on Man_5-binding, and interconversion (albeit on a slower timescale) still prevailed. However, the average T_2 value was reduced upon Man_5 binding denoting slower overall global rotational diffusion (larger τ_c) and reflecting slightly slower internal motions. This would suggest that sugar-binding further hinders the local internal motions (and the global tumbling), thus further reducing the overall flexibility of the protein. This is most pronounced in both calcium-binding loops, β-sheets 6 and 7, α-helix 1 and the N-terminus. These results support previous reports that the primary calcium binding loop and β-sheets 6 and 7 are sites of key CRD-mannose interactions, while highlighting the secondary calcium-binding loop, α-helix 1 and the N-terminus as potentially important new sites for critical lectin-glycan interactions/conformational rearrangements.

Binding of Man_9GlcNAc was again unique in that these spectra displayed severe line-broadening, which could be interpreted as further hindered dynamics in the CRD (compared with Man_5–associated CRD). It follows that the larger, higher-affinity Man_9GlcNAc could restrict/slow the motions of the CRD more than

the smaller, low affinity Man_5, and that the CRD would interconvert between a smaller number of possible conformations when this ligand was bound.

8.1.5 Ligand Binding Induces Dynamics/Conformational Changes Throughout the CRD.

Analysis of the backbone relaxation and chemical shift perturbation data shows pronounced effects in regions of the CRD distal to the proposed glycan binding site, particularly at the N-terminus. This region forms the link between the CRD and the helical neck in full length DC-SIGNR, and suggests that ligand induced signals could be propagated through the CRD, along the neck, to the cytoplasmic region of the full-length protein. Alternatively, ligand binding could alter the orientation of the CRD with respect to the neck promoting multi-valent binding by adjacent CRDs. Structural analyses of the human C-type lectin CLEC5A allude to similar possibilities that dynamic changes in the CRD, attributable to distal glycan binding, could contribute to the transmission of conformational information and signalling beyond the target binding site to the intracellular regions of the native polypeptide (Watson et al, 2011). In the case of DC-SIGNR, a receptor previously believed to be involved primarily in adhesion, evidence of signalling activity has been demonstrated in the context of respiratory syncitial virus glycoprotein binding (Johnson et al, 2012).

Changes in the CRD distal to ligand binding site along with the extent of the chemical shift perturbations mean that more residues are affected by ligand binding than can be explained by direct interaction of the protein with the oligosaccharides. This suggests that the DC-SIGNR CRD could undergo a

conformational change as a result of ligand binding and further highlights that the CRD is a highly flexible, dynamic domain that can interconvert between a number of conformations. However, the crystal structures for free and ligand-bound DC-SIGNR CRD are very similar, and do not suggest conformational rearrangement upon ligand binding. An alternative interpretation is that glycans bind to multiple binding sites or experience multivalent binding. For the homologous protein DC-SIGN, crystallography has revealed multiple binding modes for the glycans Man_2 and Man_6 (Angulo et al, 2008; Feinberg et al, 2007). However, only a single binding mode was observed in crystals of $(GlcNAc)_2Man_3$ with DC-SIGNR (Feinberg et al, 2001), and no non-linear chemical shift perturbations or broadening (which would result from multiple binding sites and/or modes) was observed after addition of the glycan fragments. The linear chemical shift perturbations also fit very well to a one-site binding model. $Man_9GlcNAc$ binding did result in broadening of signals, therefore multiple binding modes cannot be ruled out for this ligand, but taken together with the rest of the data leads to the conclusion that significant changes (both dynamic and conformational) are taking place in the CRD as a result of ligand binding in solution that cannot be sampled in the crystal structures. This inherent flexibility gives the CRD its ability to accommodate a variety of ligands and potentially transmit information throughout the protein.

8.1.6 Future Work.

Further work is needed to confirm the extent of the ligand-induced effects described above, for which solution NMR is ideally suited. On-going determination of the solution structure of both holo- and ligand-bound CRD along with saturation transfer difference experiments and further backbone relaxation measurements (particularly heteronuclear NOE experiments which will allow a more detailed analysis of the CRD dynamics through model-free analysis) will further characterise the conformational and dynamics effects of glycan binding. In addition, more work is required to confirm the apparent pH dependence of DC-SIGNR CRD-glycan binding (investigation of mannose-containing ligands is required). Complementing the NMR data with other techniques such as isothermal titration calorimetry and surface plasmon resonance could provide further information on calcium and glycan binding affinities in solution. In future, the work described here can easily and quickly be extended to investigate more ligands including disease-associated glycoproteins and potential drug molecules.

References.

Angulo J, Diaz I, Reina JJ, Tabarani G, Fieschi F, Rojo J, Nieto PM (2008) Saturation Transfer Difference (STD) NMR Spectroscopy Characterization of Dual Binding Mode of a Mannose Disaccharide to DC-SIGN. *ChemBioChem* **9:** 2225-2227

Baldisseri DM, Pelton JG, Sparks SW, Torchia DA (1991) Complete ^{1}H and ^{13}C Assignment of Lys and Leu Side-Chains of Staphylococcal Nuclease Using HCCH-COSY and HCCH-TOCSY-3D NMR-Spectroscopy. *FEBS Lett* **281:** 33-38

Baribaud F, Pohlmann S, Doms RW (2001) The role of DC-SIGN and DC-SIGNR in HIV and SIV attachment, infection, and transmission. *Virology* **286:** 1-6

Barreiro LB, Quach H, Krahenbuhl J, Khaliq S, Mohyuddin A, Mehdi SQ, Gicquel B, Neyrolles O, Quintana-Murci L (2006) DC-SIGN interacts with *Mycobacterium leprae* but sequence variation in this lectin is not associated with leprosy in the Pakistani population. *Hum Immunol* **67:** 102-107

Bashirova AA, Geijtenbeek TBH, van Duijnhoven GCF, van Vliet SJ, Eilering JBG, Martin MP, Wu L, Martin TD, Viebig N, Knolle PA, KewalRamani VN, van Kooyk Y, Carrington M (2001) A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. *J Exp Med* **193:** 671-678

Bax A (1994) Multidimensional Nuclear Magnetic Resonance Methods for Protein Studies. *Curr Opin Struc Biol* **4:** 738-744

Bax A, Clore GM, Gronenborn AM (1990) ^{1}H-^{1}H Correlation Via Isotropic Mixing of ^{13}C Magnetization, a New 3-Dimensional Approach for Assigning ^{1}H and ^{13}C Spectra of ^{13}C-Enriched Proteins. *J Magn Reson* **88:** 425-431

Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly $^{15}N/^{13}C$ enriched proteins. *J Biomol NMR* **1:** 99-104

Berthold DA, Jeisy VJ, Sasser TL, Shea JJ, Frericks HL, Shah G, Reinstra CM. (2010) Top ten tips for poducing ^{13}C ^{15}N protein in abundance. Cambridge Isotope Laboratories, Inc., Vol. Application note 15.

Bertini I, McGreevy KS, Parigi G (2012) *NMR of biomolecules : towards mechanistic systems biology*, Weinheim Chichester: Wiley-VCH ; John Wiley distributor.

Brinley FJ, Spangler SG, Mullins LJ (1975) Calcium and EDTA Fluxes in Dialyzed Squid Axons. *J Gen Physiol* **66:** 223-250

Campanero MR, Delpozo MA, Arroyo AG, Sanchezmateos P, Hernandezcaselles T, Craig A, Pulido R, Sanchezmadrid F (1993) ICAM-3 Interacts with Lfa-1 and Regulates the Lfa-1/ICAM-1 Cell-Adhesion Pathway. *J Cell Biol* **123:** 1007-1016

Caparros E, Serrano D, Puig-Kroger A, Riol L, Lasala F, Martinez I, Vidal-Vanaclocha F, Delgado R, Rodriguez-Fernandez JL, Rivas L, Corbi AL, Colmenares M (2005) Role of the C-type lectins DC-SIGN and L-SIGN in *Leishmania* interaction with host phagocytes. *Immunobiology* **210:** 185-193

Cavanagh J (2007) *Protein NMR spectroscopy : principles and practice*, 2nd edn. Burlington, Mass. ; London: Academic Press.

Chan VSF, Chan KYK, Chen YX, Poon LLM, Cheung ANY, Zheng BJ, Chan KH, Mak W, Ngan HYS, Xu XN, Screaton G, Tam PKH, Austyn JM, Chan LC, Yip SP, Peiris M, Khoo US, Lin CLS (2006) Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nat Genet* **38:** 38-46

Chen X, Tong XT, Xie YH, Wang Y, Ma JB, Gao DM, Wu HM, Chen HB (2006) Over-expression and purification of isotopically labeled recombinant ligand-binding domain of orphan nuclear receptor human B1-binding factor/human liver receptor homologue 1 for NMR studies. *PREP* **45:** 99-106

Chen YX, Chan VSF, Zheng BJ, Chan KYK, Xu XN, To LYF, Huang FP, Khoo US, Lin CLS (2007) A novel subset of putative stem/progenitor CD34(+)Oct-4(+) cells is the major target for SARS coronavirus in human lung. *J Exp Med* **204:** 2529-2536

Christensen T, Gooden DM, Kung JE, Toone EJ (2003) Additivity and the physical basis of multivalency effects: a thermodynamic investigation of the calcium EDTA interaction. *J Am Chem Soc* **125:** 7357-7366

Clore GM, Driscoll PC, Wingfield PT, Gronenborn AM (1990) Analysis of the Backbone Dynamics of Interleukin-1-β Using 2-Dimensional Inverse Detected Heteronuclear ^{15}N-^{1}H NMR-Spectroscopy. *Biochemistry* **29:** 7387-7401

Clubb RT, Thanabal V, Wagner G (1992) A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$ Chemical-Shifts in ^{15}N-^{13}C-Labeled Proteins. *J Magn Reson* **97:** 213-217

Copie V, Battles JA, Schwab JM, Torchia DA (1996) Secondary structure of β-hydroxydecanoyl thiol ester dehydrase, a 39-kDa protein, derived from Hα, Cα, Cβ and CO signal assignments and the Chemical Shift Index: comparison with the crystal structure. *J Biomol NMR* **7:** 335-340

Csizmok V, Felli IC, Tompa P, Banci L, Bertini I (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. *J Am Chem Soc* **130:** 16873-16879

Dalgleish AG, Beverley PCL, Clapham PR, Crawford DH, Greaves MF, Weiss RA (1984) The CD4 (T4) Antigen Is an Essential Component of the Receptor for the AIDS Retrovirus. *Nature* **312:** 763-767

Davis AL, Keeler J, Laue ED, Moskau D (1992) Experiments for Recording Pure-Absorption Heteronuclear Correlation Spectra Using Pulsed Field Gradients. *J Magn Reson* **98:** 207-216

Davis CW, Mattei LM, Nguyen HY, Ansarah-Sobrinho C, Doms RW, Pierson TC (2006) The location of asparagine-linked glycans on West Nile virions controls their interactions with CD209 (dendritic cell-specific ICAM-3 grabbing nonintegrin). *J Biol Chem* **281:** 37183-37194

Defougerolles AR, Springer TA (1992) Intercellular-Adhesion Molecule-3, a 3rd Adhesion Counter-Receptor for Lymphocyte Function Associated Molecule-1 on Resting Lymphocytes. *J Exp Med* **175:** 185-190

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J Biomol NMR* **6:** 277-293

Diehl C, Engstrom O, Delaine T, Hakansson M, Genheden S, Modig K, Leffler H, Ryde U, Nilsson UJ, Akke M (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc* **132:** 14577-14589

Drickamer K (1992) Engineering Galactose-Binding Activity into a C-Type Mannose-Binding Protein. *Nature* **360:** 183-186

Drickamer K (1993) Ca^{2+}-Dependent Carbohydrate-Recognition Domains in Animal Proteins. *Curr Opin Struc Biol* **3:** 393-400

Drickamer K (1997) Making a fitting choice: Common aspects of sugar-binding sites in plant and animal lectins. *Structure* **5:** 465-468

Drickamer K (1999) C-type lectin-like domains. *Curr Opin Struc Biol* **9:** 585-590

Drickamer K, Taylor ME (2005) Targeting diversity. *Nat Struct Mol Biol* **12:** 830-831

Engering A, Geijtenbeek TB, van Vliet SJ, Wijers M, van Liempt E, Demaurex N, Lanzavecchia A, Fransen J, Figdor CG, Piguet V, van Kooyk Y (2002) The dendritic cell-specific adhesion receptor DC-SIGN internalizes antigen for presentation to T cells. *J Immunol* **168:** 2118-2126

Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Formankay JD, Kay LE (1994) Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by ^{15}N NMR Relaxation. *Biochemistry* **33:** 5984-6003

Feinberg H, Castelli R, Drickamer K, Seeberger PH, Weis WI (2007) Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J Biol Chem* **282:** 4202-4209

Feinberg H, Guo Y, Mitchell DA, Drickamer K, Weis WI (2005) Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. *J Biol Chem* **280:** 1327-1335

Feinberg H, Mitchell DA, Drickamer K, Weis WI (2001) Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **294:** 2163-2166

Feinberg H, Tso CK, Taylor ME, Drickamer K, Weis WI (2009) Segmented Helical Structure of the Neck Region of the Glycan-Binding Receptor DC-SIGNR. *J Mol Biol*

Fesik SW, Eaton HL, Olejniczak ET, Zuiderweg ERP, Mcintosh LP, Dahlquist FW (1990) 2D and 3D NMR-Spectroscopy Employing ^{13}C-^{13}C Magnetization Transfer by Isotropic Mixing - Spin System-Identification in Large Proteins. *J Am Chem Soc* **112:** 886-888

Fielding L (2007) NMR methods for the determination of protein-ligand dissociation constants. *Prog Nucl Mag Res Sp* **51:** 219-242

Fogh R, Ionides J, Ulrich E, Boucher W, Vranken W, Linge JP, Habeck M, Rieping W, Bhat TN, Westbrook J, Henrick K, Gilliland G, Berman H, Thornton J, Nilges M, Markley J, Laue E (2002) The CCPN project: an interim report on a data model for the NMR community. *Nat Struct Biol* **9:** 416-418

Gabius HJ, Siebert HC, Andre S, Jimenez-Barbero J, Rudiger H (2004) Chemical biology of the sugar code. *ChemBioChem* **5:** 740-764

Geijtenbeek TB, Torensma R, van Vliet SJ, van Duijnhoven GC, Adema GJ, van Kooyk Y, Figdor CG (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* **100:** 575-585

Gramberg T, Soilleux E, Fisch T, Lalor PF, Hofmann H, Wheeldon S, Cotterill A, Wegele A, Winkler T, Adams DH, Pohlmann S (2008) Interactions of LSECtin and DC-SIGN/DC-SIGNR with viral ligands: Differential pH dependence, internalization and virion binding. *Virology* **373:** 189-201

Gramberg T, Zhu TF, Chaipan C, Marzi A, Liu HL, Wegele A, Andrus T, Hofmann H, Poehlmann S (2006) Impact of polymorphisms in the DC-SIGNR neck domain on the interaction with pathogens. *Virology* **347:** 354-363

Grzesiek S, Bax A (1992) An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *J Magn Reson* **99:** 201-207

Guntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278:** 353-378

Guo Y, Atkinson CE, Taylor ME, Drickamer K (2006) All but the shortest polymorphic forms of the viral receptor DC-SIGNR assemble into stable homo- and heterotetramers. *J Biol Chem* **281:** 16794-16798

Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K (2004) Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat Struct Mol Biol* **11:** 591-598

Guzzi C, Angulo J, Doro F, Reina JJ, Thepaut M, Fieschi F, Bernardi A, Rojo J, Nieto PM (2011) Insights into molecular recognition of $Lewis_X$ mimics by DC-SIGN using NMR and molecular modelling. *Org Biomol Chem* **9:** 7705-7712

Han S, Ba Y (2004) Determination of the concentrations of metal cations in aqueous solutions using proton NMR spectral area integration of the EDTA complexes. *J Solution Chem* **33:** 301-312

Han S, Mathias, E. V., Ba, Y. (2007) Proton NMR Determination of Mg^{2+} and Ca^{2+} Concentrations Using Tetrasodium EDTA Complexes. *Journal of physical and natural sciences* **1:** 1-5

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166:** 557-580

Harris RK, Becker ED, De Menezes SMC, Goodfellow R, Granger P (2001) NMR nomenclature. Nuclear spin properties and conventions for chemical shifts - (IUPAC recommendations 2001). *Pure Appl Chem* **73:** 1795-1818

Hiroaki H, Umetsu Y, Nabeshima Y, Hoshi M, Kohda D (2011) A simplified recipe for assigning amide NMR signals using combinatorial ^{14}N amino acid inverse-labeling. *J Struct Funct Genomics* **12:** 167-174

Ho MR, Lou YC, Wei SY, Luo SC, Lin WC, Lyu PC, Chen CP (2010) Human RegIV Protein Adopts a Typical C-Type Lectin Fold but Binds Mannan with Two Calcium-Independent Sites. *J Mol Biol* **402:** 682-695

Hwang TL, Shaka AJ (1995) Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J Magn Reson Ser A* **112:** 275-279

Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ, Thomas WD, Jr., Thackray LB, Young MD, Mason RJ, Ambrosino DM, Wentworth DE, Demartini JC, Holmes KV (2004) CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *PNAS* **101:** 15748-15753

Johnson TR, McLellan JS, Graham BS (2012) Respiratory syncytial virus glycoprotein G interacts with DC-SIGN and L-SIGN to activate ERK1 and ERK2. *J Virol* **86:** 1339-1347

Kay LE, Ikura M, Tschudin R, Bax A (1990) 3-Dimensional Triple-Resonance NMR-Spectroscopy of Isotopically Enriched Proteins. *J Magn Reson* **89:** 496-514

Kay LE, Keifer P, Saarinen T (1992) Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J Am Chem Soc* **114:** 10663-10665

Kay LE, Torchia DA, Bax A (1989) Backbone Dynamics of Proteins as Studied by ^{15}N Inverse Detected Heteronuclear NMR-Spectroscopy - Application to Staphylococcal Nuclease. *Biochemistry* **28:** 8972-8979

Keeler JH (2005) *Understanding NMR spectroscopy*, Chichester: John Wiley.

Kneller DG, Goddard TD (1993) UCSF Sparky - an NMR Display, Annotation and Assignment Tool. *University of California, San Francisco*

Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393:** 648-659

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680-685

Lameignere E, Malinovska L, Slavikova M, Duchaud E, Mitchell EP, Varrot A, Sedo O, Imberty A, Wimmerova M (2008) Structural basis for mannose recognition by a lectin from opportunistic bacteria *Burkholderia cenocepacia*. *Biochem J* **411:** 307-318

Lange NA, Speight JG (2005) *Lange's handbook of chemistry*, 16th edn. New York ; London: McGraw-Hill.

Larkin A, Imperiali B (2011) The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50:** 4411-4426

Leckband DE, Menon S, Rosenberg K, Graham SA, Taylor ME, Drickamer K (2011) Geometry and Adhesion of Extracellular Domains of DC-SIGNR Neck Length Variants Analyzed by Force-Distance Measurements. *Biochemistry* **50:** 6125-6132

Li H, Tang NLS, Chan PKS, Wang CY, Hui DSC, Luk C, Kwok R, Huang W, Sung JJY, Kong QP, Zhang YP (2008) Polymorphisms in the C-type lectin genes cluster in chromosome 19 and predisposition to severe acute respiratory syndrome coronavirus (SARS-CoV) infection. *J Med Genet* **45:** 752-758

Lian L-Y, Roberts GCK (2011) *Protein NMR spectroscopy : practical techniques and applications*, Hoboken, NJ: Wiley.

Lipari G, Szabo A (1982a) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *J Am Chem Soc* **104:** 4546-4559

Lipari G, Szabo A (1982b) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *J Am Chem Soc* **104:** 4559-4570

Lis H, Sharon N (1998) Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* **98:** 637-674

Liu ML, Mao XA, Ye CH, Huang H, Nicholson JK, Lindon JC (1998) Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. *J Magn Reson* **132:** 125-129

Mari S, Serrano-Gomez D, Canada FJ, Corbi AL, Jimenez-Barbera J (2005) 1D saturation transfer difference NMR experiments on living cells: The DC-SIGN/oligomannose interaction. *Angew Chem Int Edit* **44:** 296-298

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989a) Overcoming the Overlap Problem in the Assignment of ^{1}H-NMR Spectra of Larger Proteins by Use of 3-Dimensional Heteronuclear ^{1}H-^{15}N Hartmann-Hahn Multiple Quantum Coherence and Nuclear Overhauser Multiple Quantum Coherence Spectroscopy - Application to Interleukin-1-β. *Biochemistry* **28:** 6150-6156

Marion D, Kay LE, Sparks SW, Torchia DA, Bax A (1989b) 3-Dimensional Heteronuclear NMR of ^{15}N-Labeled Proteins. *J Am Chem Soc* **111:** 1515-1517

Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20:** 71-75

Mellman I, Fuchs R, Helenius A (1986) Acidification of the endocytic and exocytic pathways. *Annu Rev Biochem* **55:** 663-700

Miller MC, Nesmelova IV, Platt D, Klyosov A, Mayo KH (2009) The carbohydrate-binding domain on galectin-1 is more extensive for a complex glycan than for simple saccharides: implications for galectin-glycan interactions at the cell surface. *Biochem J* **421:** 211-221

Mitchell DA, Fadden AJ, Drickamer K (2001) A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR - Subunit organization and binding to multivalent ligands. *J Biol Chem* **276:** 28939-28945

Montelione GT, Lyons BA, Emerson SD, Tashiro M (1992) An Efficient Triple Resonance Experiment Using ^{13}C Isotropic Mixing for Determining Sequence-Specific Resonance Assignments of Isotopically-Enriched Proteins. *J Am Chem Soc* **114:** 10974-10975

Montoya MC, Sancho D, Bonello G, Collette Y, Langlet C, He HT, Aparicio P, Alcover A, Olive D, Sanchez-Madrid F (2002) Role of ICAM-3 in the initial interaction of T lymphocytes and APCs. *Nat Immunol* **3:** 159-168

Myszka DG, Sweet RW, Hensley P, Brigham-Burke M, Kwong PD, Hendrickson WA, Wyatt R, Sodroski J, Doyle ML (2000) Energetics of the HIV gp120-CD4 binding reaction. *PNAS* **97:** 9026-9031

Nattermann J, Ahlenstiel G, Berg T, Feldmann G, Nischalke HD, Muller T, Rockstroh J, Woitas R, Sauerbruch T, Spengler U (2006) The tandem-repeat polymorphism of the DC-SIGNR gene in HCV infection. *Journal of Viral Hepatitis* **13:** 42-46

Nesmelova IV, Ermakova E, Daragan VA, Pang M, Menendez M, Lagartera L, Solis D, Baum LG, Mayo KH (2010) Lactose Binding to Galectin-1 Modulates Structural Dynamics, Increases Conformational Entropy, and Occurs with Apparent Negative Cooperativity. *J Mol Biol* **397:** 1209-1230

Neuhaus D, Williamson MP (2000) *The nuclear Overhauser effect in structural and conformational analysis*, 2nd edn. New York ; Chichester: Wiley-VCH.

Ng KK, Drickamer K, Weis WI (1996) Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J Biol Chem* **271:** 663-674

Ng KKS, Park-Snyder S, Weis WI (1998) Ca^{2+}-dependent structural changes in C-type mannose-binding proteins. *Biochemistry* **37:** 17965-17976

Nielbo S, Thomsen JK, Graversen JH, Jensen PH, Etzerodt M, Poulsen FM, Thogersen HC (2004) Structure of the plasminogen kringle 4 binding calcium-free form of the C-type lectin-like domain of tetranectin. *Biochemistry* **43:** 8636-8643

Nokta MA, Li XD, Nichols J, Mallen M, Pou A, Asmuth D, Pollard RB (2001) Chemokine/CD4 receptor density ratios correlate with HIV replication in lymph node and peripheral blood of HIV-infected individuals. *Aids* **15:** 161-169

Nonaka M, Ma BY, Murai R, Nakamura N, Baba M, Kawasaki N, Hodohara K, Asano S, Kawasaki T (2008) Glycosylation-dependent interactions of C-type lectin DC-SIGN with colorectal tumor-associated Lewis glycans impair the function and differentiation of monocyte-derived dendritic cells. *J Immunol* **180:** 3347-3356

Onizuka T, Shimizu H, Moriwaki Y, Nakano T, Kanai S, Shimada I, Takahashi H (2012) NMR study of ligand release from asialoglycoprotein receptor under solution conditions in early endosomes. *FEBS J* **279:** 2645-2656

Paliy O, Gunasekera TS (2007) Growth of *E-coli* BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl Microbiol Biot* **73:** 1169-1172

Pavlicek J, Sopko B, Ettrich R, Kopecky V, Baumruk V, Man P, Havlicek V, Vrbacky M, Martinkova L, Kren V, Pospisil M, Bezouska K (2003) Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry* **42:** 9295-9306

Pereira-Mouries L, Almeida MJ, Ribeiro C, Peduzzi J, Barthelemy M, Milet C, Lopez E (2002) Soluble silk-like organic matrix in the nacreous layer of the bivalve *Pinctada maxima*. *Eur J Biochem* **269:** 4994-5003

Pinske C, Bonn M, Kruger S, Lindenstrauss U, Sawers RG (2011) Metabolic deficiences revealed in the biotechnologically important model bacterium *Escherichia coli* BL21(DE3). *PLoS One* **6:** e22830

Poget SF, Freund SM, Howard MJ, Bycroft M (2001) The ligand-binding loops in the tunicate C-type lectin TC14 are rigid. *Biochemistry* **40:** 10966-10972

Pohlmann S, Zhang J, Baribaud F, Chen Z, Leslie GJ, Lin G, Granelli-Piperno A, Doms RW, Rice CM, McKeating JA (2003) Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J Virol* **77:** 4070-4080

Prestegard JH (1998) New techniques in structural NMR - anisotropic interactions. *Nature Structural Biology* **5:** 517-522

Prost LR, Grim JC, Tonelli M, Kiessling LL (2012) Noncarbohydrate Glycomimetics and Glycoprotein Surrogates as DC-SIGN Antagonists and Agonists. *ACS Chem Biol* **7:** 1603-1608

Reina JJ, Diaz I, Nieto PM, Campillo NE, Paez JA, Tabarani G, Fieschi F, Rojo J (2008) Docking, synthesis, and NMR studies of mannosyl trisaccharide ligands for DC-SIGN lectin. *Org Biomol Chem* **6:** 2743-2754

Reina JJ, Sattin S, Invernizzi D, Mari S, Martinez-Prats L, Tabarani G, Fieschi F, Delgado R, Nieto PM, Rojo J, Bernardi A (2007) 1,2-mannobioside mimic: Synthesis, DC-SIGN interaction by NMR and docking, and antiviral activity. *Chemmedchem* **2:** 1030-1036

Sallum CO, Kammerer RA, Alexandrescu AT (2007) Thermodynamic and structural studies of carbohydrate binding by the agrin-G3 domain. *Biochemistry* **46:** 9541-9550

Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig KP, Kalbitzer HR (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J Biomol NMR* **39:** 275-289

Searcy DG, Greif L (1976) Apparent Thermal Destabilization of Escherichia-Coli Nucleoprotein Due to Incomplete Dialysis of EDTA. *Biochim Biophys Acta* **418:** 133-136

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44:** 213-223

Shenoy SR, Barrientos LG, Ratner DM, O'Keefe BR, Seeberger PH, Gronenborn AM, Boyd MR (2002) Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: Calorimetric and NMR characterization. *Chem Biol* **9:** 1109-1118

Simmons G, Reeves JD, Grogan CC, Vandenberghe LH, Baribaud F, Whitbeck JC, Burke E, Buchmeier MJ, Soilleux EJ, Riley JL, Doms RW, Bates P, Pohlmann S (2003) DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology* **305:** 115-123

Skelton NJ, Palmer AG, Akke M, Kordel J, Rance M, Chazin WJ (1993) Practical Aspects of 2-Dimensional Proton-Detected ^{15}N Spin Relaxation Measurements. *J Magn Reson Ser B* **102:** 253-264

Skoog DA, West DM, Holler FJ (1994) *Analytical chemistry : an introduction*, 6th edn. Philadelphia ; London: Saunders.

Snyder GA, Colonna M, Sun PD (2005a) The structure of DC-SIGNR with a portion of its repeat domain lends insights to modeling of the receptor tetramer. *J Mol Biol* **347:** 979-989

Snyder GA, Ford J, Torabi-Parizi P, Arthos JA, Schuck P, Colonna M, Sun PD (2005b) Characterization of DC-SIGN/R interaction with human immunodeficiency virus type 1 gp120 and ICAM molecules favors the receptor's role as an antigen-capturing rather than an adhesion receptor. *Journal of Virology* **79:** 4589-4598

Soilleux EJ, Barten R, Trowsdale J (2000) Cutting edge: DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. *J Immunol* **165:** 2937-2942

Soilleux EJ, Coleman N (2003) Transplacental transmission of HIV: a potential role for HIV binding lectins. *Int J Biochem Cell B* **35:** 283-287

Soilleux EJ, Morris LS, Rushbrook S, Lee B, Coleman N (2002) Expression of human immunodeficiency virus (HIV)-binding lectin DC-SIGNR: Consequences for HIV infection and immunity. *Hum Pathol* **33:** 652-659

Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189:** 113-130

Sudmeier JL, Evelhoch JL, Jonsson NBH (1980) Dependence of NMR Lineshape Analysis Upon Chemical Rates and Mechanisms - Implications for Enzyme Histidine Titrations. *J Magn Reson* **40:** 377-390

Svajger U, Anderluh M, Jeras M, Obermajer N (2010) C-type lectin DC-SIGN: An adhesion, signalling and antigen-uptake molecule that guides dendritic cells in immunity. *Cell Signal* **22:** 1397-1405

Tailleux L, Schwartz O, Herrmann JL, Pivert E, Jackson M, Amara A, Legres L, Dreher D, Nicod LP, Gluckman JC, Lagrange PH, Gicquel B, Neyrolles O (2003) DC-SIGN is the major *Mycobacterium tuberculosis* receptor on human dendritic cells. *J Exp Med* **197:** 121-127

Timpano G, Tabarani G, Anderluh M, Invernizzi D, Vasile F, Potenza D, Nieto PM, Rojo J, Fieschi F, Bernardi A (2008) Synthesis of novel DC-SIGN ligands with an α-fucosylamide anchor. *ChemBioChem* **9:** 1921-1930

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2008) BioMagResBank. *Nucleic Acids Res* **36:** D402-408

Valladeau J, Ravel O, Dezutter-Dambuyant C, Moore K, Kleijmeer M, Liu Y, Duvert-Frances V, Vincent C, Schmitt D, Davoust J, Caux C, Lebecque S, Saeland S (2000) Langerin, a novel C-type lectin specific to Langerhans cells, is an endocytic receptor that induces the formation of Birbeck granules. *Immunity* **12:** 71-81

Varki A, Etzler ME, Cummings RD, Esko JD (2009) Discovery and Classification of Glycan-Binding Proteins. In *Essentials of Glycobiology*, Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds), 2nd edn. Cold Spring Harbor (NY)

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* **59:** 687-696

Wang SF, Huang JC, Lee YM, Liu SJ, Chan YJ, Chau YP, Chong P, Chen YM (2008) DC-SIGN mediates avian H5N1 influenza virus infection *in cis* and *in trans*. *Biochem Biophys Res Commun* **373:** 561-566

Watson AA, Lebedev AA, Hall BA, Fenton-May AE, Vagin AA, Dejnirattisai W, Felce J, Mongkolsapaya J, Palma AS, Liu Y, Feizi T, Screaton GR, Murshudov GN, O'Callaghan CA (2011) Structural flexibility of the macrophage dengue virus receptor CLEC5A: implications for ligand binding and signaling. *J Biol Chem* **286:** 24208-24218

Weis WI, Drickamer K (1996) Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* **65:** 441-473

Weis WI, Drickamer K, Hendrickson WA (1992) Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* **360:** 127-134

Williamson MP (2009) Applications of the NOE in Molecular Biology. *Annu Rep NMR Spectros* **65:** 77-109

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N Chemical Shift Referencing in Biomolecular NMR. *J Biomol NMR* **6:** 135-140

Wishart DS, Sykes BD (1994) The ^{13}C Chemical Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using ^{13}C Chemical Shift Data. *J Biomol NMR* **4:** 171-180

Wishart DS, Sykes BD, Richards FM (1992) The Chemical Shift Index - a Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR-Spectroscopy. *Biochemistry* **31:** 1647-1651

Wittekind M, Mueller L (1993) Hncacb, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the α-Carbon and β-Carbon Resonances in Proteins. *J Magn Reson Ser B* **101:** 201-205

Woods RJ, Pathiaseril A, Wormald MR, Edge CJ, Dwek RA (1998) The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. *Eur J Biochem* **258:** 372-386

Wragg S, Drickamer K (1999) Identification of amino acid residues that determine pH dependence of ligand binding to the asialoglycoprotein receptor during endocytosis. *J Biol Chem* **274:** 35400-35406

Wurzburg BA, Tarchevskaya SS, Jardetzky TS (2006) Structural changes in the lectin domain of CD23, the low-affinity IgE receptor, upon calcium binding. *Structure* **14:** 1049-1058

Xu LJ, Li QQ, Ye HH, Zhang QY, Chen HC, Huang F, Chen RH, Zhou R, Zhou W, Xia PC, Chen Y, Pan C (2010) The Nine-Repeat DC-SIGNR Isoform is Associated with Increased HIV-RNA Loads and HIV Sexual Transmission. *J Clin Immunol* **30:** 402-407

Yabe R, Tateno H, Hirabayashi J (2010) Frontal affinity chromatography analysis of constructs of DC-SIGN, DC-SIGNR and LSECtin extend evidence for affinity to agalactosylated N-glycans. *FEBS J* **277:** 4010-4026

Yappert MC, DuPre DB (1997) Complexometric titrations: Competition of complexing agents in the determination of water hardness with EDTA. *J Chem Educ* **74:** 1422-1423

Yu QD, Oldring AP, Powlesland AS, Tso CKW, Yang C, Drickamer K, Taylor ME (2009) Autonomous Tetramerization Domains in the Glycan-binding Receptors DC-SIGN and DC-SIGNR. *J Mol Biol* **387:** 1075-1080

Zhu DY, Kawana-Tachikawa A, Iwamoto A, Kitamura Y (2010) Influence of polymorphism in dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin-related (DC-SIGNR) gene on HIV-1 trans-infection. *Biochem Bioph Res Co* **393:** 598-602

Appendix A: Full Chemical Shift Assignment.

Table A2: Chemical shift assignments of the holo DC-SIGNR CRD. All chemical shifts are listed in ppm and referenced to external DSS. Atoms which are not present in an amino acid are marked with an 'x' and missing assignments with a '-'.

#	amino acid	H	NH	Cα	Cβ	Cγ	Cγ2	Cδ	Cδ2	Cε	Hα	Hβ	Hβ2	Hγ	Hγ2	Hδ	Hδ2	Hε
261	A	-	-	-	-	x	x	x	x	x	-	-	x	x	x	x	x	x
262	E	-	-	62.05	-	-	x	x	x	x	-	-	-	-	-	x	x	x
263	R	8.467	122.5	58.88	32.56	-	x	-	x	x	-	-	-	-	x	-	x	x
264	L	6.632	117.8	59.56	42.51	-	x	-	-	x	-	-	x	-	x	-	-	x
265	C	9.101	121.2	51.6	40.6	x	x	x	x	x	-	-	-	x	x	x	x	x
266	R	9.296	123.3	54.98	38.47	-	x	-	x	x	5.744	2.681	1.67	1.897	x	-	x	x
267	H	9.139	126	54.42	34.68	x	x	x	x	x	-	-	-	x	x	x	x	x
268	C	-	-	-	-	x	x	x	x	x	-	-	-	x	x	x	x	x
269	P	-	-	-	-	-	x	-	x	x	-	-	-	-	x	-	-	x
270	K	9.484	124.5	-	-	-	x	-	x	-	-	-	-	-	x	-	x	-
271	D	11.53	127.3	-	-	x	x	x	x	x	-	-	-	x	x	x	x	x
272	W	8.533	120.7	57.42	29.86	x	x	x	x	x	5.313	3.661	3.268	x	x	x	x	x
273	T	9.744	118.9	62.85	71.85	x	x	x	x	x	-	-	x	x	x	x	x	x
274	F	9.061	128.9	58.15	41.13	x	x	x	x	x	5.56	3.37	3.144	x	x	x	x	x
275	F	8.752	126.6	59.31	40.95	x	x	x	x	x	3.992	2.284	2.221	x	x	x	x	x

276	Q	8.669	127.1	56.79	-	-	x	x	x	x	-	-	-	-	x	x	x	x
277	G	-	-	45.61	x	x	x	x	x	x		x	x	x	x	x	x	x
278	N	7.604	118.1	52.25	44.09	x	x	x	x	x	5.397	2.765	2.016	x	x	x	x	x
279	C	9.478	117	53.52	43.14	x	x	x	x	x	5.878	3.16	2.851	x	x	x	x	x
280	Y	9.947	121.2	57.3	41.6	x	x	x	x	x	5.86	2.979	2.826	x	x	x	x	x
281	F	8.578	125.3	55.42	40.16	x	x	x	x	x	4.275	1.242	-	x	x	x	x	x
282	M	7.732	127	53.97	31.89	33.32	x	x	x	x	4.494	1.631	1.744	2.385	x	x	x	x
283	S	7.768	119.2	59.23	62.5	x	x	x	x	x	-	-	x	x	x	x	x	x
284	N	-	-	52.54	39.81	x	x	x	x	x	5.118	2.844	2.993	x	x	x	x	x
285	S	7.255	113.4	56.87	65.33	x	x	x	x	x	-	-	x	x	x	x	x	x
286	Q	8.342	116.7	54.91	32.65	-	x	x	x	x	-	-	-	-	x	x	x	x
287	R	9.616	122.4	55.69	37.34	29.29	x	43.63	x	x	4.987	1.726	-	1.899	x	2.662	x	x
288	N	9.33	120.5	49.82	37.5	x	x	x	x	x	-	-	-	x	x	x	x	x
289	W	8.243	121	64.11	31.39	x	x	x	x	x	-	-	-	x	x	x	x	x
290	H	7.54	115.7	60.42	30.19	x	x	x	x	x	3.916	3.269	3.14	x	x	x	x	x
291	D	8.932	118.3	56.16	39.42	x	x	x	x	x	4.496	2.532	-	x	x	x	x	x
292	S	7.917	123.6	63.14	61.97	x	x	x	x	x	3.148	3.231	x	x	x	x	x	x
293	V	7.436	122.9	67.15	31.39	20.96	22.77	x	x	x	3.276	2.079	x	0.7926	-0.1229	x	x	x
294	T	7.412	114.8	65.99	82.33	x	x	x	x	x	-	-	x	x	x	x	x	x
295	A	7.866	124.6	55.32	17.88	x	x	x	x	x	4.073	1.185	x	x	x	x	x	x
296	C	7.722	112.4	55.44	36.43	x	x	x	x	x	4.879	2.809	2.696	x	x	x	x	x
297	Q	7.882	122.8	59.9	28.17	34.38	x	x	x	x	4.391	2.475	2.257	2.73	2.593	x	x	x
298	E	8.317	120	59.35	29.97	37.63	x	x	x	x	4.185	2.352	2.245	2.695	2.555	x	x	x
299	V	7.153	109.7	60.05	30.05	19.38	21.74	x	x	x	4.773	2.733	x	1.308	1.051	x	x	x
300	R	7.927	115.7	57.02	26.48	27.8	x	43.54	x	x	4.021	2.003	2.044	1.674	x	3.294	x	x

301	A	8.176	120.5	50	24.77	x	x	x	x	x	4.996	1.65	x	x	x	x	x	x
302	Q	8.124	117	54.15	33.78	34.97	x	x	x	x	5.124	2.008	2.201	2.697	2.406	x	x	x
303	L	9.109	132.9	57.83	43.02	25.6	28.31	x	-	-	5.025	1.676	x	0.2836	1.221	-	-	x
304	V	7.967	122.7	65.34	33.62	19.7	22.8	x	x	x	3.862	2.203	x	1.589	1.198	x	x	x
305	V	7.1	127.9	60.55	32.92	20.33	22.94	x	x	x	4.371	1.938	x	1.253	1.409	x	x	x
306	I	-	-	61.4	37.74	-	-	-	x	x	-	-	x	-	-	-	x	x
307	K	9.212	123.3	55.26	35.14	24.84	x	29	x	42.25	4.49	1.995	-	1.244	x	1.522	x	2.743
308	T	7.578	107	59.42	73.81	x	x	x	x	x	5.293	4.953	x	x	x	x	x	x
309	A	9.22	125.2	55.08	18.28	x	x	x	x	x	4.152	1.568	x	x	x	x	x	x
310	E	8.974	117.7	60.71	29.12	-	x	x	x	x	-	-	-	-	-	x	x	x
311	E	8.182	123.4	59.68	30.56	-	x	x	x	x	-	-	-	-	-	x	x	x
312	Q	7.757	121.3	58.92	26.16	32.03	x	x	x	x	4.327	-	-	-	x	x	x	x
313	N	8.18	117.1	56.59	38.39	x	x	x	x	x	-	-	-	x	x	x	x	x
314	F	7.701	120.5	60.58	39.25	x	x	x	x	x	4.058	2.35	2.392	x	x	x	x	x
315	L	7.992	120	57.38	42.3	-	x	-	-	x	-	-	x	-	x	-	-	x
316	Q	8.833	122.3	59.07	28.92	33.95	x	x	x	x	3.667	2.502	-	2.361	x	x	x	x
317	L	7.641	119.8	58	41.65	27.12	x	24.38	-	x	4.054	1.583	1.73	-	x	0.8994	-	x
318	Q	7.559	116.4	58.41	27.88	-	x	x	x	x	-	-	-	-	x	x	x	x
319	T	7.768	111.6	65.32	68.64	x	x	x	x	x	3.748	3.657	x	x	x	x	x	x
320	S	8.111	118.7	61.64	63.16	x	x	x	x	x	4.326	4.137	x	x	x	x	x	x
321	R	8.022	121.3	58.75	30.07	-	x	-	x	x	-	-	-	-	x	-	x	x
322	S	8.649	116	58.46	64.32	x	x	x	x	x	5.669	3.978	x	x	x	x	x	x
323	N	7.873	118.6	54.33	37.55	x	x	x	x	x	4.319	3.074	2.709	x	x	x	x	x
324	R	7.474	116.5	55.18	32.28	26.37	x	43.24	x	x	4.439	1.302	1.244	0.9352	x	2.318	1.868	x
325	F	8.679	125.1	56.84	38.52	x	x	x	x	x	5.153	3.123	3.047	x	x	x	x	x

326	S	7.432	113.8	56.5	67.53	x	x	x	x	x	5.363	3.662	x	x	x	x	x	x
327	W	9.895	124.9	57.63	34.33	x	x	x	x	x	5.356	3.626	-	x	x	x	x	x
328	M	8.389	113.9	53.42	37.98	-	x	x	x	x	-	-	-	-	-	x	x	x
329	G	9.889	112.2	48.94	x	x	x	x	x	x	3.753	x	x	x	x	x	x	x
330	L	8.354	128.3	55.07	46.58	-	x	-	-	x	-	-	x	-	x	-	-	x
331	S	8.55	120.6	58.99	68.18	x	x	x	x	x	-	-	x	x	x	x	x	x
332	D	7.966	123.8	51.9	37.7	x	x	x	x	x	4.596	2.155	-	x	x	x	x	x
333	L	6.662	119.4	57.49	44.5	26.8	x	25.15	-	x	3.709	1.487	1.346	0.7396	x	1.038	-	x
334	N	7.845	115.6	56.3	38.63	x	x	x	x	x	4.438	2.722	2.667	x	x	x	x	x
335	Q	7.754	120.7	55.25	30.55	33.88	x	x	x	x	8.493	1.914	1.86	2.2	x	x	x	x
336	E	8.075	128.2	57.16	28.49	34.32	x	x	x	x	3.934	1.677	2.281	2.485	1.896	x	x	x
337	G	10.91	120.7	45.18	x	x	x	x	x	x	3.77	x	x	x	x	x	x	x
338	T	8.5	120.9	61.88	69.57	x	x	x	x	x	4.614	4.179	x	x	x	x	x	x
339	W	8.8	128.8	57.95	29.6	x	x	x	x	x	4.261	3.23	3.006	x	x	x	x	x
340	Q	9.115	123.3	54.73	34.12	34.24	x	x	x	x	5.001	2.137	1.978	2.522	2.251	x	x	x
341	W	9.566	128.6	57.03	32.11	x	x	x	x	x	5.929	3.972	3.59	x	x	x	x	x
342	V	9.023	115.1	64.65	32.08	-	-	x	x	x	-	-	x	-	-	x	x	x
343	D	7.382	117.9	53.58	40.64	x	x	x	x	x	-	3.379	2.684	x	x	x	x	x
344	G	8.567	109	45.15	x	x	x	x	x	x	4.17	x	x	x	x	x	x	x
345	S	8.375	119.2	57.27	63.49	x	x	x	x	x	-	-	x	x	x	x	x	x
346	P	-	-	62.84	32.84	27.29	x	50.46	x	x	4.971	2.399	2.114	2.229	2.286	3.799	-	x
347	L	7.936	121.6	55.44	42.86	27.16	x	25.25	23.8	x	3.977	1.683	x	1.721	x	0.9681	-	x
348	S	8.991	127.1	56.9	63.94	x	x	x	x	x	5.178	3.141	3.032	x	x	x	x	x
349	P	-	-	-	-	-	x	-	x	x	-	-	-	-	x	-	-	x
350	S	7.736	119.9	60.81	63.09	x	x	x	x	x	4.39	4.039	x	x	x	x	x	x

351	F	8.249	120.5	57.42	39.12	x	x	x	x	x	4.589	3.524	3.85	x	x	x	x	x
352	Q	7.47	117.6	58.29	27.8	36.93	x	x	x	x	4.232	2.316	2.236	2.613	2.411	x	x	x
353	R	6.988	113.5	56.52	29.17	27.24	x	31.04	x	x	4.166	-	-	-	x	-	x	x
354	Y	6.756	118.3	58.44	34.87	x	x	x	x	x	-	-	-	x	x	x	x	x
355	W	6.444	117.9	57.86	30.69	x	x	x	x	x	4.493	3.262	2.895	x	x	x	x	x
356	N	9.92	123	53.18	39.15	x	x	x	x	x	-	3.117	-	x	x	x	x	x
357	S	114.3	8.515	60.57	63.15	x	x	x	x	x	-	-	x	x	x	x	x	x
358	G	-	-	45.72	x	x	x	x	x	x	x	x	x	x	x	x	x	x
359	E	8.258	118.9	52.81	29.62	-	x	x	x	x	-	-	-	-	-	x	x	x
360	P	-	-	-	-	-	x	-	x	x	-	-	-	-	x	-	-	x
361	N	-	-	-	-	x	x	x	x	x	-	-	-	x	x	x	x	x
362	N	-	-	54.69	-	x	x	x	x	x	-	-	-	x	x	x	x	x
363	S	7.606	114	59.55	62.3	x	x	x	x	x	4.129	-	x	x	x	x	x	x
364	G	8.588	118.1	46.91	x	x	x	x	x	x	3.971	x	x	x	x	x	x	x
365	N	7.636	114.5	55.44	36.77	x	x	x	x	x	4.241	-	-	x	x	x	x	x
366	E	7.621	123.3	56.14	31.89	26.39	x	x	x	x	4.086	1.999	-	-	-	-	x	x
367	D	7.999	125.1	51.46	41.59	x	x	x	x	x	5.498	2.787	2.656	x	x	x	x	x
368	C	8.125	116.5	59.41	49.15	x	x	x	x	x	-	-	-	x	x	x	x	x
369	A	8.102	126.5	52.07	21.38	x	x	x	x	x	5.54	1.167	x	x	x	x	x	x
370	E	9.16	117	52.4	33.39	35.2	x	x	x	x	5.67	1.706	-	2.451	2.061	x	x	x
371	F	9.382	121.1	59	41.11	x	x	x	x	x	4.046	2.282	2.231	x	x	x	x	x
372	S	8.719	117	56.42	64.66	x	x	x	x	x	-	-	x	x	x	x	x	x
373	G	8.59	114.9	47.29	x	x	x	x	x	x	-	x	x	x	x	x	x	x
374	S	-	-	58.73	64.32	x	x	x	x	x	-	-	x	x	x	x	x	x
375	G	7.575	113.6	46.61	x	x	x	x	x	x	-	x	x	x	x	x	x	-

376	W	8.369	122.4	53.02	32.59	x	x	x	x	x	6.007	2.706	2.552	x	x	x	x	x
377	N	9.26	118	51.99	42.48	x	x	x	x	x	5.571	-	-	x	x	x	x	x
378	D	9.207	120.7	52.53	42.4	x	x	x	x	x	5.374	2.757	2.012	x	x	x	x	x
379	N	9.982	125	49.34	42.07	x	x	x	x	x	5.477	2.99	-	x	x	x	x	x
380	R	9.541	122.7	56.84	30.97	-	x	-	x	x	-	-	-	-	x	-	x	x
381	C	8.57	119.9	58.64	46.17	x	x	x	x	x	-	-	-	x	x	x	x	x
382	D	8.555	114.2	54.37	40.44	x	x	x	x	x	4.231	3.065	2.837	x	x	x	x	x
383	V	7.72	121.2	63.1	32.95	21.65	22.27	x	x	x	4.207	2.497	x	0.9576	1.203	x	x	x
384	D	7.808	121.2	54.06	41.37	x	x	x	x	x	-	2.47	2.991	x	x	x	x	x
385	N	8.528	123.2	50.53	41.85	x	x	x	x	x	5.237	1.05	2.674	x	x	x	x	x
386	Y	7.188	116.4	60.3	37.98	x	x	x	x	x	-	-	-	x	x	x	x	x
387	W	-	-	-	-	x	x	x	x	x	-	-	-	x	x	x	x	x
388	I	-	-	-	-	-	-	-	x	x	-	-	x	-	-	-	-	x
389	C	-	-	-	-	x	x	x	x	x	-	-	-	x	x	x	x	x
390	K	-	-	-	-	-	x	-	x	-	-	-	-	-	x	-	x	-
391	K	-	-	-	-	-	x	-	x	-	-	-	-	-	x	-	x	-
392	P	-	-	63.43	32.56	-	x	-	x	x	-	-	-	-	x	-	-	x
393	A	7.769	123.2	52.81	18.99	x	x	x	x	x	-	-	x	x	x	x	x	x
394	A	8.757	123.5	51.16	21.05	x	x	x	x	x	-	-	x	x	x	x	x	x
395	C	-	-	-	-	x	x	x	x	x	-	-	-	x	x	x	x	x
396	F	-	-	-	-	x	x	x	x	x	-	-	-	x	x	x	x	x
397	R	-	-	56.2	31.27	-	x	-	x	x	-	-	-	-	x	-	x	x
398	D	8.302	122.5	54.59	41.34	x	x	x	x	x	5.259	2.781	2.632	x	x	x	x	x
399	E	125.8	58.07	31.52	36.69	x	x	x	x	4.214	2.121	1.972	2.253	-	x	x	x	-

Abbreviations.

+$FeCl_3$	M9 media with addition of 100 µL iron.
-$CaCl_2$	M9 media with calcium omitted.
[U-^{15}N-^{13}C]	Uniformly ^{15}N and ^{13}C labelled.
γ	Gyromagnetic ratio.
1D	One-dimensional.
2D	Two-dimensional.
3D	Three-dimensional.
B_0	Applied magnetic field.
BC	Protein sample before column purification.
BSA	Bovine serum albumin.
CRD	Carbohydrate recognition domain.
CSI	Chemical shift index.
C-type	Calcium binding lectin.
DC-SIGN	Dendritic cell specific ICAM-3 grabbing non-integrin
DC-SIGNR	Dendritic cell specific ICAM-3 grabbing non-integrin related.
dH2O	Deionized water.
d-Hepes	Deuterated hepes.
EDTA	Ethylenediaminetetraacetic acid.
FT	Flow through.
ħ	Planck's constant/2π
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus.
HSQC	Heteronuclear single quantum coherence.
I	Spin state
ICAM-3	Intracellular adhesion molecule 3.
LB	Luria-Bertani broth.
M_0	Net macroscopic magnetization.
M9	M9 minimal media.
NMR	Nuclear magnetic resonance.
NOE	Nuclear Overhauser effect.
NOESY	Nuclear Overhauser effect spectroscopy.
OD	Optical density.
P1	Pellet 1.
P2	Pellet 2.
pre-ind	Pre-induction.
R_1	Longitudinal relaxation rate.
R_2	Transverse relaxation rate.
RF	Radio frequency.
SARS-CoV	Severe acute respiratory system coronavirus.

T_1	Longitudinal relaxation time.
T_2	Transverse relaxation time.
TOCSY	Total correlation spectroscopy.
W	Wash fraction.
ω_0	Larmor frequency.
§	Section.